Journal
of Speech and Hearing
Research

June 1959 VOLUME 2 • NUMBER 2

Measures of Stuttering Adaptation 105
BRUCE QUARRINGTON

Pitch, Intensity, and Cleft Palate Voice Quality 113
DONALD A. HESS

Some Relations Between Normal Hearing for
Pure Tones and for Speech 126
JAMES F. JERGER, RAYMOND CARHART, TOM W. TILLMAN, AND
JOHN L. PETERSON

Some Factors Affecting the Spondee Threshold in
Normal-Hearing Subjects 141
TOM W. TILLMAN AND JAMES F. JERGER

Validity of the One-Frequency Screening Principle
for Public School Children 147
IRA M. VENTRY AND HAYES A. NEWBY

Dysphasic Speech Responses to Visual Word Stimuli 152
GERALD M. SIEGEL

Scaling of Abstraction Level of Single Words 161
FREDERIC L. DARLEY, DOROTHY SHERMAN, AND GERALD M. SIEGEL

Estimates of Intraglottal Pressure During Phonation 168
JAMES L. FLANAGAN

Vowel Formant Measurements 173
GORDON E. PETERSON

Unconditioned Stimulus Strength and the
Galvanic Skin Response 184
LEO G. DOERFLER AND JOAN C. KRAMER

Bibliography: Delayed Auditory Feedback 193
RICHARD ALLEN CHASE, SAMUEL SUTTON, AND DAPHNE FIRST

The American Speech and Hearing Association

OFFICERS

President
George A. Kopp, Ph.D.
Wayne State University

Executive Vice-President
Jack Matthews, Ph.D.
University of Pittsburgh

Vice-President
Miriam D. Pauls, Ph.D.
Johns Hopkins Medical Institute

Editor of the Association
Wendell Johnson, Ph.D.
University of Iowa

OFFICERS-ELECT

President-Elect
Stanley Ainsworth, Ph.D.
University of Georgia

Vice-President-Elect
Jack L. Bangs, Ph.D.
Houston Speech and Hearing Center

COUNCIL

The Officers and the following Councilors:
Jon Eisenson, Ph.D. (1959)
Ira J. Hirsh, Ph.D. (1958-60)
Ruth B. Irwin, Ph.D. (1956-59)
James F. Jerger, Ph.D. (1959-61)
Herold S. Lillywhite, Ph.D. (1958-59)
Hayes A. Newby, Ph.D. (1957-59)
Wilbert L. Pronovost, Ph.D. (1958-60)
Sylvia O. Richardson, M.D. (1957-59)
Dean E. Williams, Ph.D. (1959-61)

EXECUTIVE SECRETARY
Kenneth O. Johnson, Ph.D.

The Journal of Speech and Hearing Research

EDITOR
Dorothy Sherman, Ph.D.

ASSISTANT TO THE EDITOR
Dorothy W. Moeller

ASSISTANT EDITOR
Frederic L. Darley, Ph.D.

STATISTICAL CONSULTANT
Leonard S. Feldt, Ph.D.

ASSOCIATE EDITORS
Oliver Bloodstein, Ph.D.
Arthur S. House, Ph.D.
James F. Jerger, Ph.D.
D. E. Morley, Ph.D.
Hildred Schuell, Ph.D.
William R. Tiffany, Ph.D.
John C. Webster, Ph.D.
Joseph M. Wepman, Ph.D.

DEPARTMENT EDITORS
Ernest H. Henrikson, Ph.D.
Book Reviews
Martin F. Palmer, Sc.D.
Records

BUSINESS MANAGER
Kenneth O. Johnson, Ph.D.

APPLICATIONS FOR MEMBERSHIP SHOULD BE ADDRESSED TO THE EXECUTIVE SECRETARY

Measures of Stuttering Adaptation

BRUCE QUARRINGTON

Since its demonstration (*6*), the adaptation phenomenon has attracted the interest of many investigators of stuttering. 'Adaptation' refers to the general tendency of stuttering frequency to decrease in a positive decelerating manner as oral reading of a given prose passage is repeated. It has been shown that this effect is not confined to prose material but occurs also in repeated readings of a word list (*4*). It is also known that repetition of fixed material is not essential for frequency reduction. Subsequent sections of long prose passages when read aloud show trends of decreasing stuttering frequency (*2*, 7). It would appear, however, that reductions in frequency are most marked in repeated oral readings of prose, and this has formed the standard way of demonstrating the effect.

The length of prose passage employed seems not to affect the mean rate of adaptation shown by a group of stutterers (*12*). The time lapse following adaptation trials required to bring about spontaneous recovery of initial stuttering level has been established (*5*, *8*). The effect of differences in the social complexity or the anxiety-inducing character of the adaptation situation is not clear. Different experiments here have yielded contradictory results. Dixon (*1*) compared adaptation measures from reading situations varying mainly in their anxiety-inducing character and concluded that adaptation rate was increased with increases in situational anxiety. This increase, the author notes, is largely carried by the increased initial frequency of stuttering in the anxiety-laden situations. Shulman (*12*) compared adaptation measures collected in a standard adaptation experiment with those obtained from a situation in which the number of listeners was increased by one with each successive reading until a total of five persons including the experimenter comprised the audience for the fifth reading. His findings indicate that the complex situation depressed adaptation rate when compared to the standard situation.

The adaptation experiment has been regarded also as something of a model of the treatment situation. It has been suggested that changes in the direction of greater adaptation may reflect the degree of improvement in adjustment to stuttering (*13*). It is possible also

Bruce Quarrington (Ph.D., University of Toronto, 1953) is Associate, Department of Psychiatry, and Assistant Professor, Department of Psychology, University of Toronto. This study was made while the writer was a Fellow (1957-1958) of the Center for Advanced Studies in the Behavioral Sciences, Stanford, California.

that adaptation measures may be of value in predicting individual response to treatment. Some stutterers show a remarkably stable level of stuttering in reading after reading. One would be inclined to suspect that this inflexibility manifest in the adaptation test situation is likely to be characteristic over a longer period of time and to indicate a rather poor treatment prognosis. Marked reductions in stuttering frequency on repeated readings suggest that the stuttering of such an individual may be very responsive to increases and decreases in the affect-inducing aspects of the total situation. In such an individual any treatment resulting in decreased emotionality might be expected to result in decreased stuttering (*11*).

As far as the present writer is aware, attempts to study relations between the degree of adaptation manifest and treatment result have not yielded significant relationships. Indeed the literature gives little information as to any relationship existing between adaptation and other measures of individual differences. It is possible that one of the reasons for the variable-in-a-vacuum status of adaptation is that methods of deriving a measure of adaptation are inadequate. The present paper is an exploration of this possibility with suggestions for the development of a more useful measure of adaptation.

The most frequently employed formula for determining a measure of adaptation is one based on relative decrease from one reading to a subsequent reading (7). If the percentage of stuttering on the various readings is denoted by $p_1, p_2, p_3, \ldots, p_j$ this measure of adaptation (Ar) is given as follows:

$$Ar_n \equiv \frac{p_1 - p_n}{p_1}. \qquad (1)^1$$

Since relative decrease is usually measured from the frequency on the first reading, p_1 is given in the formula. The subsequent reading involved in this adaptation measure is often the fifth reading, but p_n is used to indicate that any subsequent reading might be employed. The subscript n is added to Ar to indicate which reading was employed.

Various workers have pointed out that measures of adaptation from this method are negatively correlated with the frequency of stuttering on the initial reading. Shulman (*12*) gives a product moment coefficient of —.64 which is a somewhat higher value than those from other studies (*10*). In the experience of the author the coefficient is usually between —.40 and —.50.

It does not appear advantageous to define adaptation by a measure that is so dependent on initial frequency. The concept of adaptation would seem of greater potential value if it were defined by a measure that is independent of initial status. It is known that stuttering frequency in most stutterers will vary considerably from day to day, and from situation to situation. There is the expectation that stuttering frequency will rise and fall with variations in the emotional state of the stutterer. With

[1]The notation of Johnson and Inness (7) has been changed here in the interest of clarity and consistency.

regard to adaptation, however, the expectation is toward stability. If adaptation is a measure of ability and interest in improving speech performance, then it should be relatively constant over short periods of time regardless of the particular emotional significance of the test situation. To employ a measure associated with initial frequency incorporates a source of unreliability in measures of adaptation.

It should be pointed out immediately that using the absolute decrease in stuttering frequency does not make adaptation independent of initial frequency. This measure, *Aa*, which has been found by the present writer to correlate with p_1 between .40 and .50, is defined as follows:

$$Aa_n \equiv p_1 - p_n. \qquad (2)$$

This problem of dependence on initial status of the usual adaptation measures was recognized by Oxtoby (*9*, *10*) who proposed the following solution. She sought to determine the most probable relative decrease for varying stuttering frequencies on the first reading. An individual's adaptation score would then be determined by dividing his observed relative decrease by the expected relative decrease for individuals having the same initial frequency. Oxtoby also introduced the idea of multiplying this ratio by a constant in order to make interpretation somewhat easier. The constant employed was the expected relative decrease when p_1 equaled 10%, which she felt was the typical initial frequency of stutterers in her experience.

The formula for this measure (*Av*) is given as:

$$Av_n \equiv V(p_1 - p_n) \qquad (3)[2]$$

$$\text{where } V = \frac{k}{\widetilde{D}_{p_1}}. \qquad (3a)$$

In this last equation k is a constant serving as an arbitrary standard. It is the experimentally obtained expected decrease $p_1 - p_n$ where $p_1 = 10\%$. The denominator $\widetilde{D}_{p_1}$ in formula (3a) is the expected decrease from *1*st to the *n*th reading for stutterers evidencing an initial percentage of p_1. This, of course, is calculated from the regression equations she derived for her groups.

Oxtoby found that this measure of adaptation (Av_5) made certain distinctions between her experimental conditions that Ar_5 did not. The method employed in *Av* strikes one as being an efficient way of bringing about independence from p_1. There is, of course, the difficulty that Oxtoby's regression equation is appropriate only for her group and that new equations must be developed for use in other studies. This, however, is but a temporary difficulty and stable regressions would probably be derived with more extensive use of the method.

There are certain inadequacies inherent in *Av* that must be pointed out. Examination of Oxtoby's data[3] indi-

[2]Original notation has been altered for purposes of presentation in this paper.

[3]All references to Oxtoby's data indicate the use of percentages of stuttering for 24 subjects on five successive readings under her 'control' conditions. These are given in Table A of her appendices (*9*).

cates that the variance of $p_1 - p_2$, $p_1 - p_3$, $p_1 - p_4$, $p_1 - p_5$ (hence also p_2, p_3, p_4, p_5) increase with p_1. This does not appear to be peculiar to these data; it has occurred in all samples of adaptation tests collected by the present writer.

The presence of heteroscedasticity in these data indicates that serious limitations are imposed on the predictive and comparative use of the derived regression equations. More immediately, it is evident that, due to the increasing variance noted, individuals with high p_1 values will tend to receive the extreme values of adaptation. If Oxtoby's data are examined for this relationship, the absolute deviations from the expected values are found to be positively correlated with p_1 to the extent of .21 to .35 ($Av_2, \ldots, Av_5$). One may also look on this as indicating that the reliability of the adaptation measures is varying inversely with initial frequency. The arguments invoked earlier in favour of a measure of adaptation independent of initial status apply here. One solution appears to be a transformation of $p_1, p_2, \ldots, p_n$, that will render variance uniform. Several transformations were tried out on Oxtoby's and other data, the most satisfactory being the angular transformation: $\phi = 2 \arcsin \sqrt{p}$ where p = percentage of stuttering (*3*, *14*, ch. 17). Henceforth $\phi_1, \phi_2, \ldots, \phi_n$ indicates the angular transformation of $p_1, p_2, \ldots, p_n$.

The measure of adaptation suggested here ($A\phi$) is based on the deviation of each particular value of ϕ_n from the predicted value $\hat{\phi}_n$ derived from the linear regression line of ϕ_n on ϕ_1. The formula is then given as:

$$A\phi_n \equiv \hat{\phi}_n - \phi_n. \qquad (4)$$

With uniformity of variance there is no need for a ratio; a simple difference suffices. It might also be noted that the correlation of ϕ_1 with $(\hat{\phi}_n - \phi_n)$ equals *zero*, which may be proved algebraically.[4]

The specific regression equations for $\phi_2, \ldots, \phi_5$ derived from Oxtoby's data are offered below:

$$\hat{\phi}_2 = .938\phi_1 - .0728$$
$$\hat{\phi}_3 = .912\phi_1 - .0525$$
$$\hat{\phi}_4 = .860\phi_1 - .0119$$
$$\hat{\phi}_5 = .849\phi_1 - .0033$$

For purposes of comparison $\hat{\phi}_2$ and $\hat{\phi}_3$ regression equations derived from another source are offered. Sixty adult stutterers applying for treatment at the Speech Clinic of the Toronto Psychiatric Hospital were required to read a prose passage ('Arthur the Young Rat') three times. This adaptation test

[4]The following proof has been offered by Dr. Leonard Feldt (x has been used instead of ϕ_1, y instead of ϕ_n):

$$r_{x(\hat{y}-y)} = \frac{\Sigma x(ax) - \Sigma xy}{N\sigma_x\sigma_{(\hat{y}-y)}}, \text{ where } \hat{y} = ax$$
$$= \frac{a\Sigma x^2 - \Sigma xy}{N\sigma_x\sigma_{(\hat{y}-y)}}$$
$$= \frac{a\sigma_x^2 - r_{xy}\sigma_x\sigma_y}{\sigma_x\sigma_{(\hat{y}-y)}}$$
$$= \frac{r_{xy}(\sigma_y/\sigma_x)\ \sigma^2_x - r_{xy}\sigma_x\sigma_y}{\sigma_x\sigma_{(\hat{y}-y)}}$$
$$= \frac{r_{xy}\sigma_y - r_{xy}\sigma_y}{\sigma_{(\hat{y}-y)}} = 0.$$

was administered during their investigation period prior to acceptance for treatment except in 12 cases for whom treatment was in progress. The instructions were essentially those employed by Oxtoby in her control conditions. The regression equations obtained were:

$$\hat{\phi}_2 = .997\phi_1 - .1041$$

$$\hat{\phi}_3 = .960\phi_1 - .1281$$

It will be noted that both these equations are more oblique and positively displaced on the dependent variable than the corresponding equations from Oxtoby's study. The displacement differences indicate that Oxtoby's subjects apparently showed greater decrements in stuttering frequency than those in the second study. These differences may be attributable largely to selection factors operating in the two samples. In Oxtoby's study the subjects were actually under treatment, whereas in the second study applicants for treatment constituted the bulk of the sample. Many of these subjects were not considered good candidates for treatment and were not accepted for therapy. Those rejected were, as a group, lower in adaptation than those accepted, although the decision to accept for treatment was made without reference to performance on the adaptation test.

The differences in the regression equations are not great despite differences in sample character (and other differences such as length of prose passage) which fosters the expectation that further studies will yield adequate normative regression equations.

Even after a way has been suggested for bringing about independence of an adaptation measure and initial frequency, there remains the task of demonstrating some increased utility of this measure over other measures. Although some evidence will be offered later indicating advantages of $A\phi$, the major burden of proof has yet to be taken up. At this point there appears to be some merit in turning to another problem in connection with adaptation measures to see what help, if any, the measure $A\phi$ offers.

The problem is which subsequent reading (p_n) is to be employed in determining a measure of adaptation. As mentioned earlier, the fifth reading is most frequently employed. The reason appears to be that this is a critical balance point of reward in terms of observed reduction in stuttering frequency and effort involved in administering the test. Beyond the fifth reading, group adaptation curves plotted in percentages flatten out and the test situation becomes markedly more tedious for both subject and observer. Another reason for choice of the fifth reading may be that Ar_5 distributions are usually normal, whereas Ar_2, Ar_3, Ar_4 distributions are often positively skewed. Distributions from measure $Ar_6, \ldots,$ Ar_n usually show negative skewness. Neither are compelling reasons for the selections of any particular reading as p_n. The only reason that would be compelling is the demonstration that adaptation measures based on a particular reading yield stronger relationships with other variables of interest (for example, outcome of treatment). Lacking information of this sort, it is necessary to assume that any reading subsequent to the initial one offers as good a

TABLE 1. Comparison of intertrial correlations of two adaptation measures (Ar, $A\phi$).

		$A\phi_4$	$A\phi_3$	$A\phi_2$	
		.914	.796	.824	$A\phi_5$
Ar_4	.664		.777	.768	$A\phi_4$
Ar_3	.770	.696		.716	$A\phi_3$
Ar_2	.580	.638	.659		
	Ar_5	Ar_4	Ar_3		

basis for measuring adaptation as any other. It is possible, of course, temporarily to settle on a particular reading to be employed by studying the intercorrelations of adaptation measures based on different choices of p_n and selecting that reading having the highest set of correlations. Presumably this is the measure with the highest loading of an adaptation factor, or may be taken as such until demonstrated otherwise.[5]

The intercorrelations of $A\phi_2, \ldots,$ $A\phi_5$ and $Ar_2, \ldots, Ar_5$ were determined using Oxtoby's data. These are given in Table 1 with the $A\phi$ intercorrelations in the upper-right half of the matrix and the Ar intercorrelations in the lower-left position.

Inspection of the $A\phi$ intercorrelations shows a simple structure with the correlations, with one exception, steadily declining from $A\phi_5$ to $A\phi_2$. One would conclude, considering only five readings, that the fifth offers the best basis for deriving adaptation measures. When the Ar intercorrelations are considered, structure is ambiguous but aparently indicates the superiority of the third reading. The lack of simplicity in the pattern of correlations, however, suggests the intrusion of extraneous factors in Ar measures. This is also suggested by the lower coefficients of Ar intercorrelations when compared to those of $A\phi$ measures. Without exception $A\phi$ correlations exceed those of their corresponding Ar correlations. These observations must be considered quite tentative, of course, in view of the small number of cases (24) on which the correlations are based.

An indication of the possible value of $A\phi$ will be offered in connection with the matter of retest reliability. It has been pointed out that initial frequency of stuttering may be expected to vary considerably in a stutterer from one testing to another. This suggests the $A\phi$ measures should be superior to Ar measures when evaluated in terms of test-retest correlations.

A study of 14 adult stutterers was conducted. Two adaptation tests were given to each subject from one to eight days apart. Two different 180-word passages, comparable in word difficulty and frequency of initial sounds, were employed. One passage was administered by the speech therapist in charge of stutterers' pretreatment investigation, while the second passage was administered by a member of the psychology staff with whom the stutterers were not so familiar.

The intention was to compare the performance of Ar and $A\phi$ measures under two test conditions which varied in time, prose material, and particularly in the affect-inducing character of the test situation. Under these conditions considerable intraindividual variation among the subjects in level of stuttering

[5]At the present time combinations of more than two readings are not considered as a means of deriving adaptation measures, although possibilities here are clearly worth investigation.

frequency would be expected. The product moment correlation coefficient of initial frequencies was in fact .59. Turning to adaptation measures, the test-retest correlations for these 14 subjects were $A\phi_3 = .69$ and $Ar_3 = .42$. This does not constitute a demonstration of adequate test-retest reliability, nor was this the intention. The results do suggest, however, that studies which control conditions in the interest of showing the reliability of adaptation measures may find $A\phi$ measures of greater value than Ar measures.

Summary

Relationships of adaptation measures to other variables of individual differences among stutterers are lacking in the literature. It is conjectured that this may be due to inadequate measures of adaptation and accordingly several existing measures are examined in this light. A likely source of error appears to be the dependence of adaptation measures on frequency of stuttering on the initial reading of the adaptation experiment.

An adaptation measure ($A\phi$) independent of initial frequency is proposed that involves the angular transformation of stuttering frequencies. Deviations from the expected level of stuttering are used as adaptation scores. The expected scores are determined by reference to experimentally obtained linear regressions of the amount of stuttering of subsequent readings on the amount of stuttering of initial reading. Some examples of regression equations are offered.

Under experimental conditions accentuating intraindividual variability of initial stuttering frequencies an $A\phi$ measure was shown to have a higher test-retest correlation than a measure based on simple relative decrease.

References

1. Dixon, Carmen C., Stuttering adaptation in relation to assumed level of anxiety. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 12.
2. Donohue, Irene R., Stuttering adaptation during three hours of continuous oral reading. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 17.
3. Eisenhart, C., Hastay, M. W., and Wallis, W. A., *Selected Techniques of Statistical Analysis.* New York: McGraw-Hill, 1947. Ch. 7, 16.
4. Golub, A., The cumulative effect of constant and varying reading material on stuttering adaptation. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 13.
5. Jamison, D. J., Spontaneous recovery of the stuttering response as a function of the time following adaptation. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 14.
6. Johnson, W., and Knott, J. R., Studies in the psychology of stuttering: I. The distribution of moments of stuttering in successive readings of the same material. *J. Speech Dis.*, 2, 1937, 17-19.
7. Johnson, W., and Inness, Marjorie, Studies in the psychology of stuttering: XIII. A statistical analysis of the adaptation and consistency effects in relation to stuttering. *J. Speech Dis.*, 4, 1939, 79-86.
8. Jones, E. L., Explorations of experimental extinction and spontaneous recovery in stuttering. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 11.
9. Oxtoby, Eloise T., A quantitative study of certain phenomena related to expectancy of stuttering. Ph.D. thesis, Univ. of Iowa, 1946.

10. Oxtoby, Eloise T., Frequency of stuttering in relation to induced modifications following expectancy of stuttering. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 10.
11. Quarrington, B., Cyclical variation in stuttering frequency and some related forms of variation. *Canad. J. Psychol.*, 10, 1956, 179-184.
12. Shulman, E., Factors influencing the variability of stuttering. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 9.
13. Van Riper, C., and Hull, Catherine J., The quantitative measurement of the effect of certain situations on stuttering. In W. Johnson (Ed.), *Stuttering in Children and Adults.* Minneapolis: Univ. of Minnesota Press, 1955. Ch. 8.
14. Walker, Helen M., and Lev, J., *Statistical Inference.* New York: Henry Holt, 1953.

Pitch, Intensity, and Cleft Palate Voice Quality

DONALD A. HESS

Hypernasality, breathiness, and hoarseness, according to McDonald and Baker (*16*), are voice quality characteristics which may be discernible in the voices of cleft palate speakers. Berry and Eisenson (*2*) include harshness as a possible vocal concomitant of cleft palate speech. Since these four terms have been used in some speech correction texts (*6, 12, 28*) to describe most of the types of perceptual voice-quality aberration, it would appear that the manner in which cleft palate voice may deviate from normal is highly diversified.

That a significant degree of nasality associated with cleft palate may result from morphologic involvements in the speech structures, particularly the velum, is well recognized, although a number of functional factors have been discussed (*13, 16*). Hoarseness among cleft palate speakers has been explained as a result of chronically inflamed and swollen tissues (2, p. 330; *16*), habitual glottal articulation (*22*, p. 165), and ventricular voice (*2*, p. 329). Breathiness in cleft palate speech may be a vocal consequence of 'faulty phonation' (*16*), while harshness is believed by other writers (2, p. 326) to be the result of lingual, pharyngeal, and laryngeal muscular imbalances associated with glottal articulation. Van Hattum (*26*) discovered that his cleft palate subjects used a greater percentage of their available air supply during phonation and had lower vital capacities than his normal subjects. It seems reasonable that some degree of deviant voice quality may result from the manner in which cleft speakers learn to use their speaking voices.

Vocal pitch and intensity have been found to relate functionally to problems of breathiness (*6*), harshness (*6*), and hoarseness (*19, 20, 30, 33*). Opinion and evidence from certain writers (*5, 29, 32*) suggest that vocal weakness and nasality are functionally related. Statements (*7, 8, 9, 10, 11, 14, 18, 21*) about the effect of pitch on the perception of nasality are less conclusive, however, and one objective study of functionally nasal speakers (*24*) failed to discover a relationship between pitch level and perceived nasality.

Kantner (*14*) stated that the perception of nasality must remain the major

Donald A. Hess (D.Ed., Pennsylvania State University, 1955) is Professor of Education and Psychology and Director, Speech and Hearing Clinic, State Teachers College, Indiana, Pennsylvania. This article is based on a D.Ed. dissertation completed under the direction of Dr. Eugene T. McDonald. A portion of it is adapted from a paper presented at the 1957 convention of the American Speech and Hearing Association, Cincinnati.

basis for discussing it. His observation probably holds true for other types of voice quality disturbance as well. With special reference to nasality, Johnson, Darley, and Spriestersbach (*13*, p. 71) have questioned the wisdom of making inferences about cleft palate voice from research findings based on subjects with functional etiologies.

On the basis of the literature, it thus appeared that an objective investigation of the effects of pitch and intensity on perceptible cleft palate voice quality was indicated. The problem, then, was to study the effects of vocal pitch and intensity on the degree to which nasality, breathiness, harshness, and hoarseness are perceived in the voices of cleft palate speakers.

Procedure

Preliminary Considerations and Studies. The experimental plan included (a) instructing cleft palate speakers to vocalize at different pitch and intensity levels, (b) recording their vocalizations, and (c) obtaining auditor judgments with respect to nasality, breathiness, harshness, and hoarseness of the recorded voices.

Subject Delimitations. To avoid possible extraneous variation undesirable to the purposes of the study, the subjects were limited to male cleft palate speakers who had undergone voice change. Other requirements for their selection were that they (a) possess borderline or higher intelligence, (b) pass a sweep-frequency screening, hearing test (20 db) in the better ear for 500-4000 cps, and (c) perform successfully in the final test conditions. No criterion for degree of vocal defectiveness was used, nor were the subjects limited to any particular type of palatal restoration.

Experimental Speech Sample. The disadvantages of a connected speech sample to obtain judgments of voice quality have been discussed by Sherman (*23*). Moreover, it was necessary to use a speech sample which could be more easily controlled for conditions of pitch and intensity. For these reasons, isolated, sustained vowels were used. Six vowels—[i], [u], [e], [o], [æ], and [ɑ]—were selected to represent the high, intermediate, and low positions of the front and back vowels. In a pilot study it was found that voice quality could be most effectively judged when these vowels were sustained for about 2.5 seconds.

Experimental Pitch and Intensity Levels. To ensure that the findings and conclusions of the study might be more directly applicable to clinical problems, it was felt that the test pitch levels for each subject should be based on his own habitual pitch. It was hoped that the present study could avoid the atypical vocal results that the use of standard pitch levels might involve. On the other hand, standard intensity levels which could be attained by all subjects were desired. Since Cobb and Lierle (*4*) had observed that cleft palate speakers are deficient in 'pitch variation and intensity' during connected speech, it was necessary to discover how capably the subjects could alter their vocal intensities at different pitch levels.

In a pilot study 20 male cleft palate speakers were tested to determine their ability to phonate the six vowels

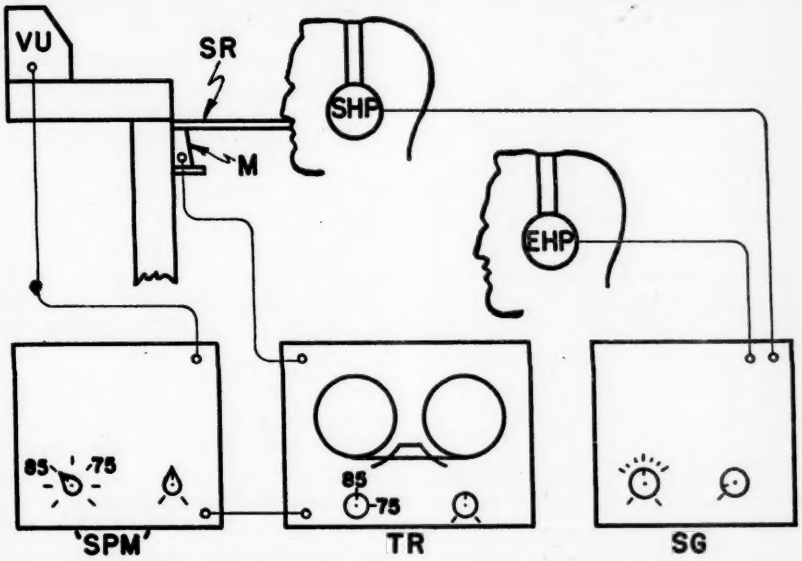

FIGURE 1. Simplified diagram of the equipment used to test the subjects and obtain the recorded vowel phonations. Key: SR, spacing rod used to control mouth-to-microphone distance at six inches; VU, VU meter; M, microphone; 'SPM,' amplifier rigged to act as a sound pressure meter; TR, tape recorder; SG, signal generator; SHP, subject's headphone; EHP, experimenter's headphone.

through ranges of 90 to 210 cps and 65 to 100 db[1] SPL (*re*: .0002 dyne/cm^2) at a mouth-to-microphone distance of 6 in. It was found that a majority of the subjects could phonate the test vowels within a pitch range between habitual pitch (in cps) and a higher pitch level 1.4 times the numerical cps value for habitual pitch. Within these pitch ranges they could phonate the vowels between 75 and 85 db SPL. These extremes were thus selected as the experimental pitch and intensity levels. Since Steer and Tiffin (*25*) had found that the superior speakers they studied had an average range in syllabic power of approximately 20 db, with a standard deviation of 4.4 db, it was felt that the selected intensity levels would represent relatively soft and loud voices.

Estimate of Habitual Pitch. To determine whether he could reliably judge habitual pitch, the experimenter instructed 21 male cleft palate speakers to (a) repeat a standard test sentence five times and (b) discuss a subject like school, home town, or hobby for a short period of time. These two types of speech samples were tape recorded. From the recordings, the habitual pitch of each speaker was estimated by matching the experimenter's subjective

[1]These intensity levels were measured with an H. H. Scott sound level meter, Type 410-A, with C setting.

judgment of habitual pitch with a pure tone generated from a direct-reading signal generator. After a lapse of one week and with a new random order of speaker voices, a second estimate of habitual pitch was made from each type of sample. The resultant Pearson *r*s were .98 and .97, respectively, for the repeated estimates of habitual pitch from these two types of speech samples; the *t* tests of mean differences in estimate from judgment to rejudgment failed to show significant differences. It was thus felt that considerable confidence could be placed in his ability to estimate habitual pitch reliably.

Overview of Final Recording and Judgment Procedure. Fifteen male cleft palate speakers who had undergone voice change were instructed to phonate each of six vowels—[i], [u], [e], [o], [æ], and [ɑ]—for 2.5 sec at each of two pitch levels (habitual pitch in cps and a pitch level 1.4 times higher) and two intensity levels (75 db SPL and 85 db SPL). These vowel phonations were tape recorded. In all, 24 vowel phonations were tape recorded for each subject. The 360 recorded vowel phonations were randomized for pitch levels, intensity levels, vowels, and subjects. They were played back for auditor judgments of severity of nasality, breathiness, harshness, and hoarseness.

Equipment. Figure 1 is a simplified diagram of the equipment used to obtain the recordings of the test vowels phonated by the subjects at the different experimental pitch and intensity levels. A steel spacing rod, ⅜ in. in diameter, was used to control mouth-to-microphone distance at 6 in. Intensity of phonation was controlled by an amplifier unit which had been calibrated with an H. H. Scott sound level meter, Type 410-A (C setting), to enable it to act as a sound pressure meter. For fine monitoring, a VU meter was used. To help the subjects maintain a proper pitch level while phonating the vowels, a Hewlett Packard audio oscillator, Model 200 CR, with monaural headphones for subject and experimenter, was employed. Acting as a signal generator, it supplied pure tones that represented the test pitch levels for each subject.

Because it was necessary to obtain the recordings of the subjects in their homes, a Pentron tape recorder, Model TR 4, with crystal microphone and tape speed of 7.5 in. per second, was used as the recording instrument. Its frequency response from 50 to 9500 cps is advertised as ±3 db, excluding the speaker. Frequency response tests indicated that the instrument was adequate for recording the subjects' voices. However, the playback for judgments of voice quality required the use of a Magnecorder. Model PT6-A, operating at a tape speed of 7.5 in. per second; a Magnecorder amplifier, Model PT6-J; and an Altec Lansing, Model 604C, 15-in. coaxial speaker.

Subjects. Fifteen subjects were used in the recording procedure. They ranged in age from 15 to 42 years, with a mean age of 19.1 years. Six subjects had postoperative palates, six were wearing prostheses, and three had neither type of palatal restoration although they were in the process of being fitted with prostheses. The subjects had undergone 0 to 200 hours of individual therapy; the average therapy time was 80 hours. They ranged in habitual pitch

from 96 to 163 cps, with a mean habitual pitch of 133.1 cps. This average habitual pitch is very similar to that reported by Fairbanks (*6*, pp. 168,169) for adult male superior speakers. However, since all potential subjects could not master the experimental pitch and intensity conditions, and since a larger group of subjects had been found to possess a somewhat lower mean habitual pitch in a preliminary study, the average habitual pitch of the 15 subjects finally employed does not necessarily represent that of adult male cleft palate speakers in general.

Recording Procedure. The habitual pitch of each subject was determined by using the method described by Fairbanks (*6*, pp. 166, 167). Since some subjects were poor readers, a standard test sentence which was to be repeated over and over was substituted for the reading passage. The experimenter's subjective estimates of habitual pitch for two trials were numerically translated into cps by matching the vowel tones with pure tones from the direct-reading signal generator. The average of these two numerical values was regarded as habitual pitch, while the higher pitch level was determined simply by multiplying habitual pitch in cps by 1.4. In their repetitions of the test sentence the subjects had a tendency to drop their average pitch levels. When this was the case, the subject was trained further until the appropriate pitch level was discovered.

Prior to the tape recording of each vowel phonation, each subject was trained to sustain the test vowel for 2.5 sec at the required pitch and intensity level. Pitch level was controlled by requiring the subject to match his pitch with the pure tone heard in his headphone. The subject monitored intensity by keeping the VU meter steady if possible and within ± 2 db of the required level.

All test vowels for both intensity levels were first phonated by the subjects at the higher pitch level, then at the lower pitch level. An important reason prompted this particular procedure. It had been found that some subjects could be trained to phonate the vowels at a given pitch level only with considerable difficulty. Once they succeeded, they could continue to produce that pitch on successive trials, even with the required changes in intensity. When they were required to change pitch on successive trials, however, they had to be retrained, sometimes with great consumption of time.

For the 12 phonations within each pitch level, the six test vowels were rotated with the two intensity levels. The signal generator was set to emit a barely audible continuous pure tone representing the test pitch level. For training the subjects and recording each vowel phonation, the sound-pressure-meter amplifier was set at the appropriate intensity setting, and the tape recorder volume control was set at the appropriate 'best recording level.'[2] Tone controls of the sound pressure meter and tape recorder remained constantly at flat setting.

There were three exceptions to this procedure: (a) Four subjects who had moderately severe to severe monaural hearing loss did not wear headphones but were trained to produce the re-

[2]Best recording level' was defined as that maximum gain setting which allowed all test vowels to be recorded at a given intensity level without distortion.

quired pitch level by the experimenter. (b) Two subjects could phonate [u] with an intensity no higher than 81 or 82 db. In each case the tape-recorder gain control was advanced to permit a best recording. (c) Three subjects could not be initially trained to produce the higher pitch level and were tested first through the habitual pitch level.

The time required to train the subjects and obtain the recorded vowel phonations varied from about 40 min to 1 hr, 45 min.

In the preliminary test of the frequency response of the recording microphone, amplifier, and recording head, as well as in the playback through the Magnecorder PT-6, the greatest deviation (about 5 db) in an otherwise markedly flat overall frequency response was for 200 cps. With an average habitual pitch of 133.1 cps, the subjects phonated at their higher pitch level with fundamentals in this approximate energy area. Unfortunately a correction had not been made for this possible source of error.

Preparation of Recordings. A total of 360 separate vowel phonations (24 for each of 15 subjects) was recorded. The tapes for each were cut and coded for subjects, vowels, pitch levels, and intensity levels. They were then reassembled in a randomized order, with an additional provision that no two consecutive vowels would be from the same subject. Announcements for the order of presentation were inserted in the tape before each vowel phonation, and a 45-in. length of blank tape followed each to allow 6 sec for rating voice quality.

To obtain sample vowel phonations for training the judges prior to each rating session, the experimenter selected those which, in his opinion, represented extremes in the continuum for each voice quality attribute being judged. Thus two extremely nasal, two extremely breathy, two extremely harsh, and two extremely hoarse samples were chosen, along with several vowel phonations which appeared to be free from any disturbing voice quality characteristic. Finally, four vowel phonations were picked at random. For each voice quality characteristic to be rated, the appropriate vowel-phonation dubs were assembled in this order: (a) extremely defective, (b) normal or nondefective, (c) extremely defective, (d) normal or nondefective, and (e) randomly selected. Thus, there were five vowel-phonation dubs for training the judges prior to the actual judgments of each voice-quality attribute: nasality, breathiness, harshness, and hoarseness. The 20 vowel-phonation dubs were placed at the beginning of reel I.

Voice Quality Judgments. In the selection of the four voice quality characteristics to be judged, Fairbanks (*6*) was the authority followed. It was reasoned that clear definitions which include descriptions of the perceptual characteristics of each type of deviant voice quality must be formulated if meaningful judgments are expected. Therefore, in each rating session the voice quality attribute to be rated was discussed in detail with the judges, and operational definitions were given. For nasality, the critical characteristic to be judged was degree of perceived nasal resonance (*6*, p. 203). Denasality was defined as 'no nasality,' to be rated as *zero* on the scale. Breathiness was de-

TABLE 1. Summary of analysis of variance to test effects of pitch levels, intensity levels, and vowels on summed ratings of nasality.

Source	*df*	*ms*	*F**	$F_{.05}$†
Pitch Levels (P)	1	336.400	32.74	4.60
Intensity Levels (I)	1	88.010	12.80	4.60
Vowels (V)	5	162.932	10.81	2.37
Subjects (S)	14	156.517		
PI	1	.100	.03	4.60
PV	5	7.246	1.13	2.37
PS	14	10.275		
IV	5	6.404	1.43	2.37
IS	14	6.874		
VS	70	15.075		
PIV	5	4.240	.71	2.37
PIS	14	3.737		
PVS	70	6.407		
IVS	70	4.482		
PIVS	70	5.963		
Total	359			

* F ratios: ms_P/ms_{PS}; ms_I/ms_{IS}; ms_V/ms_{VS}; ms_{PI}/ms_{PIS}; ms_{PV}/ms_{PVS}; ms_{IV}/ms_{IVS}; ms_{PIV}/ms_{PIVS}.

† $F_{.05}$ is the tabled value for the nearest given *df*.

fined after West, Kennedy and Carr (*31*, p. 417) as 'a vocal defect in which a portion of the expired air is not resolved into tone.' For this attribute, the judges were cautioned to listen for the audible rushing of air (*6*, p. 209) in making their judgments. Harshness was defined as 'characterized by a noisy, raspy, unmusical tone. When low pitched it frequently is described as "guttural"; when high pitched the term "strident" sometimes is applied' (*6*, p. 213). Fairbanks' (*6*, p. 216) definition of hoarseness also was used: 'From both acoustical and causal points of view hoarseness combines the features of breathiness and harshness.'

Four graduate students with extensive experience in judging voice quality were employed as the judges. For the judgments of each voice quality characteristic, a seven-point equal-appearing-intervals rating scale was employed. The scale in each case ranged from *zero*, representing no voice quality disorder, to *six*, representing an ex-

TABLE 2. Mean severity ratings for each voice quality characteristic at the various experimenta pitch and intensity levels.

Experimental Conditions	*Mean Severity Ratings*			
	Nasality	*Breathiness*	*Harshness*	*Hoarseness*
Pitch Levels				
HP	3.84	3.04	2.49	1.71
1.4 HP	3.36	2.96	2.18	1.38
Intensity Levels				
75 db SPL	3.73	3.77	2.21	1.81
85 db SPL	3.48	2.23	2.46	1.28
(General Means)	3.60	3.00	2.33	1.55

TABLE 3. Summary of analysis of variance to test effects of pitch levels, intensity levels, and vowels on summed ratings of breathiness.

Source	*df*	*ms*	*F**	$F_{.05}$†
Pitch Levels (P)	1	11.380	.50	4.60
Intensity Levels (I)	1	3397.880	165.15	4.60
Vowels (V)	5	139.558	12.84	2.37
Subjects (S)	14	108.517		
PI	1	25.600	2.89	4.60
PV	5	18.470	2.27	2.37
PS	14	22.610		
IV	5	66.718	7.03	2.37
IS	14	20.574		
VS	70	10.871		
PIV	5	15.000	2.14	2.37
PIS	14	8.856		
PVS	70	8.125		
IVS	70	9.493		
PIVS	70	7.006		
Total	359			

*F ratios: ms_P/ms_{PS}; ms_I/ms_{IS}; ms_V/ms_{VS}; ms_{PI}/ms_{PIS}; ms_{PV}/ms_{PVS}; ms_{IV}/ms_{IVS}; ms_{PIV}/ms_{PIVS}.

†$F_{.05}$ is the tabled value for the nearest given *df*.

treme voice quality disorder. For all judgments, discrete numbers were required.

To obtain the judgments, the recorded phonations were played back in a sound-treated room at a standard level of comfortable loudness. Four separate rating sessions were held. In each, one voice quality attribute was judged, and 40 consecutive vowel phonations were re-rated for purposes of a reliability check. Rest periods of 10 minutes were given after each reel of 120 vowel phonations. The order in which the voice quality characteristics were judged in the separate rating ses-

TABLE 4. Summary of analysis of variance to test effects of pitch levels, intensity levels, and vowels on summed ratings of harshness.

Source	*df*	*ms*	*F**	$F_{.05}$†
Pitch Levels (P)	1	138.130	7.09	4.60
Intensity Levels (I)	1	99.220	4.93	4.60
Vowels (V)	5	206.962	19.91	2.37
Subjects (S)	14	77.288		
PI	1	.810	.11	4.60
PV	5	15.524	2.20	2.37
PS	14	19.494		
IV	5	9.560	1.43	2.37
IS	14	20.119		
VS	70	10.394		
PIV	5	5.640	.75	2.37
PIS	14	7.076		
PVS	70	7.058		
IVS	70	6.701		
PIVS	70	7.553		
Total	359			

* F ratios: ms_P/ms_{PS}; ms_I/ms_{IS}; ms_V/ms_{VS}; ms_{PI}/ms_{PIS}; ms_{PV}/ms_{PVS}; ms_{IV}/ms_{IVS}; ms_{PIV}/ms_{PIVS}.

†$F_{.05}$ is the tabled value for the nearest given *df*.

TABLE 5. Summary of analysis of variance to test effects of pitch levels, intensity levels, and vowels on summed ratings of hoarseness.

Source	*df*	*ms*	*F**	$F_{.05}$†
Pitch Levels (P)	1	164.020	11.56	4.60
Intensity Levels (I)	1	390.620	34.46	4.60
Vowels (V)	5	60.310	11.51	2.37
Subjects (S)	14	33.642		
PI	1	.230	.04	4.60
PV	5	4.332	1.39	2.37
PS	14	14.186		
IV	5	22.878	5.30	2.37
IS	14	11.334		
VS	70	5.239		
PIV	5	1.158	.27	2.37
PIS	14	6.159		
PVS	70	3.121		
IVS	70	4.315		
PIVS	70	4.236		
Total	359			

* *F* ratios: ms_P/ms_{PS}; ms_I/ms_{IS}; ms_V/ms_{VS}; ms_{PI}/ms_{PIS}; ms_{PV}/ms_{PVS}; ms_{IV}/ms_{IVS}; ms_{PIV}/ms_{PIVS}.

†$F_{.05}$ is the tabled value for the nearest given *df*.

sions was as follows: nasality, breathiness, harshness, and hoarseness. With about two hours required for each rating session, the total procedure for obtaining auditor judgments entailed approximately eight hours of time.

Results

Reliability of Criterion Measures. The criterion measure was the sum of the four ratings by four judges for the phonation of one vowel by one subject. Reliability was evaluated for each set of 40 repeated measures (summed ratings) and the corresponding initial measures, by means of the Pearson *r* procedure along with the *t* test of the difference between means. The obtained *r*s for nasality, breathiness, harshness, and hoarseness measures were .74, .88, .85, and .75, respectively. None of the differences between means were significant at the 5% level. Although the reliability coefficients were somewhat lower than would be desirable, it was assumed that the summed ratings would be sufficiently reliable for the purposes of this study.

Method of Analysis. The obtained measures for each of the four voice qualities were analyzed by analysis of variance (*15*, pp. 254, 257) with four factors: pitch levels, intensity levels, vowels, and subjects in a 2 x 2 x 6 x 15 classification. There were 360 measures for each analysis.

Effect of Pitch and Intensity on Perceived Nasality. Pitch levels, intensity levels, and vowels[3] exerted a significant[4] effect on the degree of nasality perceived (see Table 1). The observed effects of pitch and intensity are reported in Table 2. The differences between means indicate less nasality at the higher pitch level than at the habitual

[3]Specific effects of vowels on perception of voice quality are not reported in this article.

[4]In this and in subsequent analyses, the 5% level was required for significance.

pitch level and less nasality at the more intense level, 85 db SPL, than at the less intense level, 75 db SPL.

Effect of Pitch and Intensity on Perceived Breathiness. Degree of perceived breathiness varied significantly with intensity levels and with vowels; the effect of pitch, however, was not significant (see Table 3). Perceived breathiness was markedly less severe at the more intense vocal level than at the less intense level (see Table 2). The significant vowel-intensity interaction is the result, apparently, of relatively small differences (see Table 2) for [u] and for [o] between mean severity measures obtained for the two intensity levels.

Effect of Pitch and Intensity on Perceived Harshness. Severity of perceived harshness varied significantly as an effect of pitch levels, of intensity levels, and of vowels (see Table 4). Harshness is on the average less severe for the higher pitch level than for habitual pitch level and less severe for the lower intensity level than for the higher intensity level (see Table 2).

Effect of Pitch and Intensity on Perceived Hoarseness. Pitch levels, intensity levels, and vowels exerted a significant effect on the degree of hoarseness perceived by the judges (see Table 5). Also significant is the interaction of intensity levels with vowels. An examination of the data for vowels revealed that, while the subjects were heard as considerably less hoarse at 85 db SPL while phonating the vowels [ɑ], [æ], and [e], this effect did not hold as markedly for the vowels [u], [i], and [o]. There was thus consistency of direction of trend and the test of the main effect of intensity levels has a meaningful interpretation regardless of the significant interaction. The mean severity ratings for hoarseness were lower at the higher pitch level than at habitual pitch level and lower for the higher intensity level than for the lower intensity level (see Table 2).

Summary of Results. Examination of Tables 1, 3, 4, and 5 shows that the interactions of pitch with intensity are nonsignificant for each of the four quality disorders studied. The effects of intensity on the perception of each of the four different voice quality attributes may thus be assumed to be consistent from one pitch level to another (and *vice versa*).

In summary, at the higher pitch level (1.4 times the numerical value in cps for habitual pitch) nasality, harshness, and hoarseness are significantly less severe than at habitual pitch level. The pitch level effect for breathiness is nonsignificant. For nasality, breathiness, and hoarseness the severity is significantly less for the higher intensity than for the lower intensity. Harshness is significantly more severe at the higher intensity.

Discussion and Conclusions

Vowels as Speech Samples. Although, as previously explained, standard intensity levels for the vowel phonations were desired in the present study, several experimental observations led the writer to question their use with different vowels. As previously explained, two subjects could not phonate [u] at 85 db SPL. Many of the subjects phonated [i] only with difficulty at this intensity level, whereas they had little difficulty with the other vowels, particularly [ɑ] and [æ]. At the lower (75 db SPL) level, however,

they tended to exhibit greater vocal control in their phonation of [i] and [u], the very vowels with which they had experienced difficulty at the more intense phonation level. Black (*3*) reported lower average intensities and peak power for the high vowels than for the low vowels in connected speech. The present observations on phonations of different vowels would suggest that high vowels may typically have lower intensities for the reason that they cannot be phonated at higher intensities with the ease that characterizes the phonation of low vowels, while the converse would hold true at lower intensities. Thus, it would seem reasonable that the use of standard intensity levels for the phonation of different vowels might not result in findings that are truly representative of some of the vowels as they are normally phonated in connected speech. An objection also might be raised to the use of sustained vowels, since they certainly are not typical of connected speech.

Assuming negligible error in the manner by which sound pressure level was controlled in the present study, different vowels also apparently have different 'best recording levels.' In similar future studies a potential source of error might be avoided by correction for these differences, for example, by the use of a 'best recording level' for each particular test vowel.

Pitch and Intensity. The present finding that nasality is perceived in lesser degree at higher pitch levels is in agreement with Froeschels (*9*). It would disagree with earlier opinions of Gray and Wise (*10*) and Holmes (*11*), as well as the findings of Sherman and Goodwin (*24*), who, however, used functionally nasal subjects and a connected speech sample in their study.

Weiss's (*29*) finding of a moderate but significant correlation (—.57) between oral sound pressure levels and judged severity of nasality led him to conclude that the louder a nasal voice is, the less nasal it is heard to be. The present findings tend to corroborate his conclusions, as well as statements by Williamson (*32*) and Cotton (*5*) in this regard.

Fairbanks (*6*, p. 210) states that breathiness may be caused by the use of a pitch level lower than natural pitch. Since the natural pitch of the subjects in the present study was not determined, the present findings for the negligible effect of pitch on perceived breathiness cannot be used to evaluate this statement. The observation that breathiness is perceptibly reduced at greater vocal intensity would strongly corroborate Fairbanks' (*6*, p. 212) statement that a very loud voice cannot at the same time be breathy.

Harshness may result from 'strain to produce adequate loudness, particularly at low pitches,' according to Fairbanks (*6*, p. 214). Since the highest mean ratings for severity of harshness were assigned to the louder vowel phonations at hibitual pitch, and since all the more loudly phonated vowels were perceived on the average as more harsh, the present results tend to confirm his statement.

That hoarseness tends to be perceptually reduced at pitch levels higher than habitual pitch is apparently true for functionally hoarse speakers (*33*), persons whose hoarseness is related to organic conditions such as laryngeal

contact ulcers (*19*), and adult male cleft palate speakers. However, as different authors (*1, 9, 31,* p. 141) attest, the constant use of unnaturally high pitch levels also may involve undesirable vocal consequences. The finding that the subjects were perceived as less hoarse while phonating more intensely is not what one might predict from statements by certain writers (*19, 20, 30*). It would seem pertinent to point out that this particular experimental result might be fallacious if unreservedly applied to speaking behavior. Most persons, as occasions may demand, can vocalize with tonal clarity at much greater intensities than they normally use. It is, however, a recognized fact that continued use of loud voice may lead to serious vocal disorders.

Implications for Therapy. The present results, along with the observation of Cobb and Lierle (*4*) that vocal weakness and restricted pitch variation characterized their cleft palate subjects, lead this writer to conclude that speech therapists may profitably experiment with judiciously elevated pitch and increased intensity as therapeutic adjuncts to enable the cleft palate speaker to reduce at least some of his perceptually disturbing voice quality, particularly nasality.

Several cautions should be kept in mind in applying these conclusions to speech therapy. Findings reported by other investigators (*17,* 27) suggest that the nasality perceived in cleft palate speech may be partly a function of articulatory proficiency. Moreover, Van Hattum (27) concluded that nasality judged from isolated, sustained vowels is an inadequate predictor of nasality judged from connected speech. Further, it is recognized that, to some degree, the judges in the present study may have been required to rate a perceptual artificiality caused by the subjects' prolongation of vowels at intensity levels which may not be typical of their normal production in connected speech. The procedure by which intensity was controlled may well have caused the subjects to depart from their normal manner of vocal production in their attempts to keep the VU meter needle steady. Finally, the number of experimental pitch and intensity levels was quite limited.

Van Riper (*28,* p. 472) has described a procedure for teaching cleft palate speakers to use low pitches and low intensities while blending vowels as a preliminary to articulation therapy. However, he gives no rationale for this therapeutic technique. If improvement in voice quality is his intended goal, the present results might raise a question on the routine use of this clinical procedure.

Summary

The effects of pitch and intensity level on perceived cleft palate voice quality were studied. Fifteen adult male cleft palate subjects were instructed to phonate six test vowels at two pitch levels (habitual pitch and a higher pitch) and two intensity levels. Tape recordings of these vowel phonations were played back for auditor judgments of nasality, breathiness, harshness, and hoarseness.

According to the obtained results, nasality, harshness, and hoarseness are less severe at the higher pitch level than at the habitual pitch level. Breathiness is unaffected by pitch level. At the more intense level of phonation there

is less nasality, less breathiness, more harshness, and less hoarseness than at the less intense level.

References

1. Allen, Barbara, and Peterson, G. E., Laryngeal inflammation in a case of falsetto. *J. Speech Dis.*, 7, 1942, 175-178.
2. Berry, Mildred F., and Eisenson, J., *Speech Disorders*. New York: Appleton-Century-Crofts, 1956.
3. Black, J. W., Natural frequency, duration, and intensity of vowels in reading. *J. Speech Hearing Dis.*, 14, 1949, 216-221.
4. Cobb, Lois H., and Lierle, D. M., An analysis of the speech difficulties of 56 cleft palate and harelip cases. *Arch. Speech*, 1, 1936, 217-230.
5. Cotton, J. C., A study of certain phoniatric resonance phenomena. *J. Speech Dis.*, 5, 1940, 289-293.
6. Fairbanks, G., *Voice and Articulation Drillbook*. New York: Harper, 1940.
7. Fitz-Gibbon, J. J., Cleft palate. *Proc. Amer. Soc. Study Dis. Speech*, Vol. 4, New York, 1933, 52-55.
8. Froeschels, E. (Ed.), *Twentieth Century Speech and Voice Correction*. New York: Philosophical Library, 1948.
9. Froeschels, E., Should the speech therapist be a voice therapist? *J. Speech Hearing Dis.*, 13, 1948, 346-350.
10. Gray, G. W., and Wise, C. M., *The Bases of Speech*. New York: Harper, 1934.
11. Holmes, F. L. D., The qualities of the voice. *Quart. J. Speech*, 18, 1932, 249-255.
12. Johnson, W., Brown, S. J., Curtis, J. F., Edney, C. W., Keaster, Jacqueline, *Speech Handicapped School Children*. New York: Harper, 1956.
13. Johnson, W., Darley, F. L., and Spriestersbach, D. C., *Diagnostic Manual in Speech Correction*. New York: Harper, 1952.
14. Kantner, C. E., Diagnosis and prognosis in cleft palate speech. *J. Speech Hearing Dis.*, 13, 1948, 211-222.
15. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin, 1953.
16. McDonald, E. T., and Baker, H. K., Cleft palate speech: an integration of research and clinical observation. *J. Speech Hearing Dis.*, 16, 1951, 9-20.
17. McWilliams, Betty Jane, Some factors in the intelligibility of cleft-palate speech. *J. Speech Hearing Dis.*, 19, 1954, 524-527.
18. O'Neill, J. M., and Weaver, A. T., *The Elements of Speech*. New York: Longmans, Green and Co., 1926.
19. Peacher, Georgiana, Contact ulcer of the larynx: part IV. a clinical study of vocal re-education. *J. Speech Dis.*, 12, 1947, 179-190.
20. Russell, G. O., Etiology of follicular pharyngitis, catarrhal laryngitis, so-called clergyman's throat, and singer's nodes. *J. Speech Dis.*, 1, 1936, 113-122.
21. Seely, H. F., and Hackett, W. A., *Experiences in Speaking*. New York: Scott, Foresman and Co., 1940.
22. Seth, G., and Guthrie, D., *Speech in Childhood, Its Development and Disorders*. London: Oxford Univ. Press, 1935.
23. Sherman, Dorothy, The merits of backward playing of connected speech in the scaling of voice quality disorders. *J. Speech Hearing Dis.*, 19, 1954, 312-321.
24. Sherman, Dorothy, and Goodwin, F., Pitch level and nasality. *J. Speech Hearing Dis.*, 19, 1954, 423-428.
25. Steer, M. D., and Tiffin, J., A photographic study of the use of intensity by superior speakers. *Speech Monogr.*, 1, 1934, 72-78.
26. Van Hattum, R. J., A study of air use by cleft palate and normal speaking subjects. Unpublished M. S. Thesis, Pennsylvania State Univ., 1952.
27. Van Hattum, R. J., Articulation and nasality in cleft palate speakers. *J. Speech Hearing Res.*, 1, 1958, 383-387.
28. Van Riper, C. R., *Speech Correction: Principles and Methods*. New York: Prentice-Hall, 1954.
29. Weiss, A. I., Oral and nasal sound pressure levels as related to judged severity of nasality. Unpublished Ph.D. dissertation, Purdue Univ., 1954.
30. West, R., The function of the speech pathologists in studying cases of dysphonia. *J. Speech Dis.*, 3, 1938, 81-84.
31. West, R., Kennedy, L., and Carr, A., *The Rehabilitation of Speech*. New York: Harper, 1947.
32. Williamson, A. B., Diagnosis and treatment of eighty-four cases of nasality. *Quart. J. Speech*, 30, 1944, 471-479.
33. Williamson, A. B., Diagnosis and treatment of seventy-two cases of hoarse voice. *Quart. J. Speech*, 31, 1945, 189-202.

Some Relations Between Normal Hearing for Pure Tones and for Speech

JAMES F. JERGER

RAYMOND CARHART

TOM W. TILLMAN

JOHN L. PETERSON

Serious confusion exists as to the difference between the sound pressure levels which characterize the normal threshold for a 1000-cps pure tone and the normal threshold for speech. Estimates as to the size of this discrepancy disagree by as much as 10 db. In consequence, there is need for clarification of the situation since the relationship between these two thresholds must be taken into account in establishing coordinated audiometric standards.

The problem at hand is quickly apparent when one compares the norms as presently described by the American Standards for Audiometers with the findings of various studies reported in the literature. The American Standard Specification for Audiometers for General Diagnostic Purposes (*1*) defines the normal threshold sound pressure level for 1000 cps as 16.5 db (SPL referred to 0.0002 microbar in a National Bureau of Standards Coupler 9-A), for the Western Electric type 705-A earphone. The American Standard Specification for Speech Audiometers lists 22 db (SPL referred to 0.0002 microbar in an American Standard type-1 Coupler) as the norm for speech. Indeed, the Specification explicitly states that, '. . . the purpose of this requirement [22 db SPL] is to set the 0 hearing loss for speech at a level about *6 db* [authors' italics] above the "normal" [sic] threshold for a pure tone of 1000 cps as defined in American Standard Audiometers for General Diagnostic Purposes . . .' (2, p. 9). The Specification further implies that the 6-db difference is recognized as an approximation which may require revision.

Sporadic evidence (*3, 4, 6, 9, 11*) suggests that the difference in question is considerably greater than 6 db. Davis (*6*), for example, reports the average of the thresholds for 500, 1000, and 2000 cps as 9 db (SPL) at the same time that he gives thresholds for various speech materials which range from 22 db for spondees, through 26 db for

James F. Jerger (Ph.D., Northwestern University, 1954) is Assistant Professor of Audiology; Raymond Carhart (Ph.D., Northwestern University, 1936) is Professor of Audiology; Tom W. Tillman (M.A., University of Oklahoma, 1955) is Research Assistant; John L. Peterson (M.A., Western Reserve University, 1952) is Research Assistant, all at Northwestern University. This study was supported by funds provided under Contract AF41(657)-185 with the USAF School of Aviation Medicine, Randolph Field, Texas.

sentence material and digits, to 33 db for PB words as spoken by Rush Hughes. Lightfoot, Carhart, and Gaeth (*11*) observed a 16.5-db difference between the threshold intensities for 1000 cps and for spondee words exhibited by 31 otologically normal subjects. The most definitive finding to date, however, is derived from the 1954 Wisconsin State Fair Survey (*9*). Here, the difference between the averages of the median threshold values for a 1000-cps pure tone and for spondee words, reported for all ears in the 'selected normal group' was 15 db. Discrepancies of relatively similar size characterize the results obtained in the Survey for samples (by decades) of the general population, although the data as presented must be converted to sound pressure levels before the fact is fully apparent. Most recently, there is the indirect evidence to be derived from Corso's (*4, 3*) studies of normals' thresholds for pure tones and for CID Auditory Test W-2. On relatively large groups of normal listeners, he obtained threshold SPLs of about 5 db for a 1000-cps pure tone and approximately 19 db for the W-2 spondee recording. These data imply that the difference between thresholds is on the order of 14 db.

Recent evidence (*3, 4,* 7, *10*), particularly the work of Dadson and King (*5*) in England, has made it imperative to ask whether the present American norms for pure tones are correct. There is pressure in many quarters to alter these norms as a requisite step toward establishment of an international standard for pure-tone audiometers. This situation, among other things, intensifies the need to define the relation between thresholds for pure tones and for speech.

It is now clearly apparent from evidence such as that obtained in the Wisconsin State Fair Survey (*9*) that young adults yield better thresholds than a less select group of 'normal' listeners derived from the population at large. Moreover, the present American norm for speech (22 db SPL) apparently represents performance of selected young adults (the so-called 'laboratory ear'). The American norm for pure tones, on the other hand, is based on the responses of a less restricted sampling (*12*). Here the 'man on the street with "normal" hearing' served as the referant. Thus, the two current American Standards seem to be in disagreement as to the category of normalcy on which they rest. If so, the situation must be rectified by relating both sets of audiometric specifications to the same criterion population.

The first step toward such a unification of audiometric standards is to define the difference in intensity between thresholds for pure tones and for speech. Particularly pressing in this respect, since the sound pressure level of a speech signal is defined in terms of the sound pressure level of an equivalent 1000-cps pure tone (*2*), is the need to know the relative acuity for speech and for 1000 cps.

The present investigation was undertaken to explore the latter need. The experimental problem was to ascertain the physical discrepancy between thresholds for a 1000-cps pure tone and for speech. The specific procedure was to measure both thresholds in the same normal-hearing subjects. Recorded spondees were employed as the speech

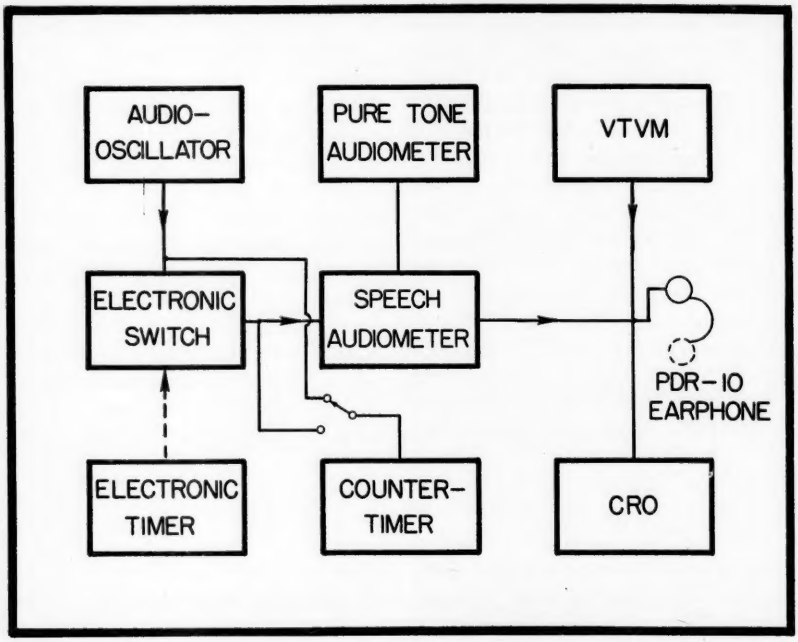

FIGURE 1. Block diagram of apparatus used to measure both pure-tone and spondee thresholds. The core of the equipment is a commercially-available speech audiometer feeding a PDR-10 earphone.

material. Variables suspected of having a critical influence on the relationship between the two thresholds were examined. These variables included sophistication of the listener, effect of practice, method of threshold determination, order of test administration, sex, ear, and familiarity with test vocabulary. Two groups of 10 subjects each were compared in examining the first two variables, while a third group of 96 subjects was used in a counterbalanced routine designed to study the remaining five variables.

Apparatus

Figure 1 shows a simplified block diagram of the experimental apparatus used to measure both pure-tone and spondee thresholds. The core of the equipment consisted of a commercially-available speech audiometer (Grason-Stadler, type 162) feeding a PDR-10 earphone mounted in an MX41/AR cushion. All spondee thresholds were obtained by playing either recorded list E or recorded list F of CID Auditory Test W-1 through this speech audiometer.

Two separate pure-tone sources were fed through the speech audiometer to the same PDR-10 earphone. One source consisted of an audio-oscillator (General Radio, type 1304A) whose output was controlled by an electronic switch (Grason-Stadler, type 829). The

electronic switch was, in turn, triggered by an electronic timer (Grason-Stadler, type 471), to produce the desired temporal pattern of the pure-tone stimulus. Stability of both the oscillator frequency and the duration of each short tone passed by the switch was assured by continuous monitoring of both oscillator and switch outputs with a counter-timer (Berkeley, type 5500). This pure-tone source was used to measure the 1000-cps threshold by a rigidly defined psychophysical procedure described below. The second pure-tone source was an ordinary pure-tone diagnostic audiometer (ADC, model 53C). This source was used to measure the 1000-cps threshold by the 'clinical' method which was studied.

The electrical signal across the earphone was monitored with a vacuum-tube voltmeter (Hewlett-Packard, Model 400C) and a cathode-ray oscillograph (Dumont, Model 304-A).

The acoustic output of the apparatus was calibrated by means of an ASA Type-1 coupler, calibrated condenser microphone (Western Electric, type 640AA), cathode follower (ADC, D5153) and vacuum-tube voltmeter (Hewlett-Packard, Model 400-A). All thresholds, both pure-tone and speech, are thus subsequently reported as the sound pressure level, referred to 0.0002 microbar, developed in an ASA Type-1 coupler.

The sound pressure level developed by the earphone at 1000 cps was measured daily throughout the course of the experiment. The maximum variation in output over the five-month period during which subjects were tested was 1.4 db, and the day-to-day variation exhibited no systematic trend over time. Pressure levels developed at octave frequencies from 125 to 8000 cps were measured weekly over the five-month period, and demonstrated equivalent stability.

Prior to the experiment, precautions were taken to insure that the speech audiometer conformed to all requirements listed in the American Standard Specifications for Speech Audiometers (*2*) with particular reference to tests of overall acoustic fidelity. The apparatus equalled or exceeded all listed specifications.

Threshold Measurement Techniques

In order to evaluate the possible effect of threshold measurement technique on the relationship between pure-tone and speech threshold, two separate and distinct methods for defining each type of threshold were employed. One, the 'up-and-down' method (*8*) was characterized by a relatively rigidly-defined set of operations. The other, hereafter called the 'clinical' method, was as nearly as possible a duplication of the procedures commonly employed by audiologists within the context of the clinical situation.

Up-and-Down Method. The up-and-down method was selected in preference to any of the three classical psychophysical methods or their variants because it permitted a relatively identical stimulus temporal pattern and measurement procedure for both pure-tone and speech thresholds. The essence of the method is that the intensity level of each successive stimulus is determined by the subject's response to the previous stimulus. If the subject does not respond at a particular intensity, the next level of the stimulus is

increased by a predetermined fixed amount. On the other hand, if the subject does respond at a particular intensity, the level of the next stimulus is decreased by the same amount. This simple rule, raising the intensity when the subject does not respond and lowering it when he does respond, is followed throughout the course of a predetermined number of stimulus presentations. The result is a series of response measures which oscillate about the threshold intensity. From these data the intensity level corresponding to 50% response may be determined by suitable arithmetic computation (*8*).

In the present experiment the intensity levels of successive stimuli were altered in 2-db steps over a series of 36 presentations. For the spondee threshold, the block of successive stimuli consisted of the 36 spondee words recorded as lists E or F of CID Auditory Test W-1. For the 1000-cps threshold, the stimulus sequence was a train of 36 short tones recurring at 6-sec intervals. Each short tone had a rise-decay time of 50 msec, and a duration, at maximum amplitude, of 500 msec. The 6-sec repetition rate for the pure tones was selected to match the rate at which the spondee words recur on the W-1 recordings. Thus, insofar as thresholds obtained by the up-and-down method are concerned, the experimental procedure was virtually identical for both pure-tone and spondee thresholds. The only difference was that for the spondee threshold each stimulus was a word which the subject repeated, either correctly or incorrectly, while for the 1000-cps threshold each stimulus was a pure tone to which the subject either did, or did not, respond.

Clinical Method. Devising a satisfactory analogue to so-called 'clinical audiometric technique' proved to be one of the more difficult problems encountered in this investigation. Numerous sources were consulted in an attempt to find some common denominator epitomizing the basic operations to be followed in the clinical measurement of an auditory threshold. The relatively small number of even cursorily described procedures found in the literature were characterized by a certain lack of agreement on some relevant particulars (for example, the number of times a stimulus is presented at a given level, whether a pure-tone stimulus is briefly turned on or briefly turned off, whether the threshold criterion is 100% response, 50% response, 0% response, or some intermediate value).

Briefly, it seemed apparent that, in order to introduce some minimal degree of objectivity into the 'clinical' thresholds, it was necessary for the authors to devise their own method. In so doing, the attempt was made to follow, as closely as possible, the counsel of experienced clinicians. In the final analysis the authors are able to justify the 'clinical' procedures ultimately employed on the sole basis that they seemed reasonable to them.

The clinical procedure adopted for the measurement of the spondee threshold corresponded closely to the method described by Newby (*13*, pp. 119-120) for the W-1 records. Two or three words were initially presented at a level 20 to 30 db above the estimated threshold level. Successive blocks of two or three words were then progressively attenuated in 10-db steps until a level was reached at which two consecutive

words were repeated incorrectly. At this point, the tester simply 'jumped around,' in no set order, from level to level in 2-db steps, presenting exactly four words per level. The spondee threshold was recorded as the lowest intensity at which the subject repeated two out of four words correctly. In the event that the subject repeated three out of four words correctly at one level, and only one out of four correctly at the next lower level, threshold was recorded as the intensity yielding three out of four responses. This problem seldom arose in actual practice.

The clinical procedure employed to measure the 1000-cps threshold closely resembled the 'ascending' technique described in the 1951 revision of the 'Manual for School Hearing Conservation Programs' (*14*, p. 15) prepared by the Committee on Conservation of Hearing of the American Academy of Ophthalmology and Otolaryngology. By means of the interruptor switch on the clinical audiometer, a brief tone was first presented at a level 30 db above the estimated threshold in order to familiarize the subject with the test signal. The tester then descended in 10-db steps, presenting one brief tone at each level, until the subject failed to respond. The tester next ascended in 5-db steps, presenting one brief tone at each level, until a response occurred. He then decreased the intensity by 10 or 15 db and again ascended in 5-db steps until another response occurred. This procedure was followed until the subject had responded three times at the same level. Threshold was thus defined as the lowest intensity at which the subject responded three times in ascending runs, using 5-db steps, and presenting just one stimulus per step. No attempt was made to control the duration of each tonal presentation other than to instruct the tester to keep the presentation brief. In practice, the tones were about 1 to 2 sec long.

Instructions to each subject tested by either the clinical method or the up-and-down method were as follows:

> The purpose of this study is to measure your threshold for tones and for words. Two test runs will be conducted, one using tones and one using words.
>
> During the tone test, you will hear a short burst of sound followed by intervals of silence. Each tone will be quite short. Some will be easy to hear. Others will be very faint. Whenever you hear one of these tones, no matter how faint it is, press the button. Since the tones will be very faint, it is necessary that you listen very carefully.
>
> When I test for your word threshold, you will hear a man's voice saying two-syllable words, such as 'wigwam,' 'therefore' or similar words. Each word will be preceded by the phrase 'Say the word.' It is only necessary for you to repeat the two-syllable word, not the phrase. Some of the words will be easy to hear. Others will be very faint. Whenever you hear a word, no matter how faint it is, repeat it out loud. You will have to listen carefully since the words will be very faint. At the beginning of each word test, you will hear several sentences of identifying information which you do not need to repeat.
>
> There is no set order for the two test runs. You will be told at the beginning of each test whether it is to be a word test or a tone test.
>
> After I have placed the phones over your ears, it is extremely important that you do not move them in any way until the tests are completed. I will tell you when you can take them off. Any questions?

Subjects

All subjects were audiometrically screened at a hearing level (hearing loss dial setting) of 10 db *re* USPHS norm

at octave intervals from 125 to 8000 cps, and at 1500 and 3000 cps, in order to insure that each subject had relatively normal acuity in both ears.

As already mentioned, three groups of subjects were employed. One group consisted of 10 sophisticated listeners. These individuals were selected from the staff and the graduate student population at the Audiological Laboratory. Each was highly experienced at the task of listening for very faint signals, and all were relatively familiar with the CID revised spondee word lists. They represented essentially 'laboratory ears.' Five subjects were male, the other five, female. They ranged in age from 20 to 31 years.

The second group was composed of 10 undergraduate students selected on the basis of having had no previous experience as listeners in auditory tests of any kind. Nine subjects were female, the other one, male. They ranged in age from 18 to 25 years.

The third group included 96 subjects, three for each of 32 separate experimental conditions. As was true of the second group, these 96 subjects were selected from the undergraduate population at Northwestern University and met the requirement that they had not had prior experience with any auditory tests. No subject was accepted who reported any history of either ear pathology or excessive noise exposure.

Prior Exposure to Spondee Words. Subjects differed in their familiarity with the spondee words. The 10 sophisticated listeners were well acquainted with these materials, while all other subjects initially were not. However, half of the 96 persons were exposed to the spondees immediately prior to the measurement of their thresholds for the words. To this end, these 48 subjects were given additional instructions as follows:

> Before the word test, I will read a series of 36 two-syllable words at a level which is easy for you to hear. You are to repeat each word. These words are the same words which you will later hear in the word test; however, they will be in a different order. Since the purpose of this initial reading of the words is to make you familiar with the words, please listen carefully.

Procedure and Results

Preliminary Study. An initial investigation was conducted to determine the effects of sophistication in auditory tests, and of practice upon the relationship between thresholds for pure tones and for speech. It was for this purpose that the first two groups of 10 subjects were formed.

Six thresholds were obtained for each subject in a single experimental session. Four of these thresholds were for pure tones of 500, 1000, 1500 and 2000 cps, respectively. The last two were spondee thresholds obtained separately with W-1 list E and with W-1 list F. All six thresholds, pure-tone and speech, were measured by the up-and-down method previously described using 2-db steps of attenuation. Subsequent comparison of the two groups yielded information on the effect of familiarity with audiological procedures.

Subjects in the sophisticated group underwent the foregoing procedure twice. The first run gave these subjects experience on the specific tasks involved. The results obtained during this session supplied the base of reference against which to estimate the effect of a practice session on thresholds

TABLE 1. Mean pure-tone and spondee (W-1) threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for sophisticated (N=10) and unsophisticated (N=10) listeners.

Threshold	*Sophisticated Group*		*Unsophisticated Group*
	Session 1	*Session 2*	
Pure Tone			
500 cps	12.4	12.1	10.3
1000 cps	6.8	7.3	5.7
1500 cps	7.4	6.7	7.7
2000 cps	6.6	7.1	8.8
Spondee			
First Test	17.0	16.8	24.3
Second Test	16.7	16.6	21.4
Difference	0.3	0.2	2.9
Mean	16.8	16.7	22.8
Mean Spondee Minus 1000 cps	10.0	9.4	17.1

measured by the up-and-down method.

Table 1 summarizes the findings of the preliminary investigation. The mean pure-tone threshold sound-pressure levels obtained for the sophisticated group in the first experimental session (practice session) were essentially equivalent to those obtained in the second session (test session). This was true also for the mean spondee threshold. In addition, when the mean spondee thresholds for this group are considered, no difference is noted between the first and second thresholds obtained in either the practice session or the test session.

Further inspection of Table 1 reveals that the pure-tone threshold sound-pressure levels are essentially equivalent for both sophisticated and unsophisticated listeners. When the spondee threshold sound-pressure levels for the two groups are compared, however, it becomes apparent that the sophisticated listeners yielded much lower thresholds than the unsophisticated listeners. Furthermore, unsophisticated listeners improved an average of about 3 db from the first threshold to the second, while sophisticated listeners did not.

It seemed possible that this 3-db improvement between the first and second spondee thresholds in the unsophisticated group could be due to the fact that the subjects gained knowledge of the test vocabulary during the first test. The same reasoning would account for the lack of improvement on the second spondee threshold in the sophisticated group. In other words, the sophisticated subjects probably already had the optimum degree of familiarity with the words prior to the administration of the first spondee test so that further exposure to test vocabulary had no demonstrable effect on threshold. If this interpretation is correct, one would expect the unsophisticated group to show further improvement on successive retests with spondee words until they reached a terminal level of familiarity, equivalent to that of the sophisticated group.

The foregoing results indicated the necessity of controlling, in the main experiment, both the general factor of subject sophistication in auditory tests

and the specific factor of prior familiarity with the CID revised Harvard spondee words. The decision was made to restrict the main experiment to audiologically naïve listeners because such persons are more similar to the relatively unsophisticated population encountered in clinical situations. Moreover, it was deemed particularly important, since such persons showed instability of threshold with successive exposures to test material, to explore more fully the effect of familiarity with test procedure upon threshold levels. Information on this point is essential if the variability of threshold due to retesting is to be taken into account in specifying the criterion population for audiometric norms.

Main Experiment. A five-factor design was undertaken to assess the effects on thresholds for spondee words and for a 1000-cps tone of: (a) threshold measurement technique, (b) order of test administration, (c) sex, (d) ear, and (e) prior knowledge of test vocabulary. This investigation was undertaken with the third group of subjects, 96 young adults having normal hearing but lacking prior experience with auditory tests. In order to control systematically the five factors under consideration, the group was appropriately divided. Specifically, 48 subjects were tested by the up-and-down method, the other 48 by the clinical method. Within each subgroup, the spondee test was administered first to half of the subjects, the pure-tone test first to the rest. Half of the subjects were female, and the remainder, male. The right ear was tested 50% of the time, and the left ear the remaining 50%. The 36 spondees were read to half of the subjects prior

TABLE 2. Mean 1000-cps and spondee threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for each method of measurement (N=96).

Threshold	*Method*	
	Up-and-Down	*Clinical*
1000 cps	8.1	9.8
Spondee	21.6	21.7

TABLE 3. Mean 1000-cps and spondee threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for each sex (N=96).

Threshold	*Sex*	
	Male	*Female*
1000 cps	9.3	8.6
Spondee	21.3	22.0

TABLE 4. Mean 1000-cps and spondee threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for each ear (N=96).

Threshold	*Ear*	
	Right	*Left*
1000 cps	8.3	9.6
Spondee	21.7	21.5

TABLE 5. Mean 1000-cps and spondee threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for each test order (N=96).

Threshold	*Test Order*	
	Spondee First	*1000 cps First*
1000 cps	9.7	8.2
Spondee	22.3	21.0

TABLE 6. Mean 1000-cps and spondee threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) according to prior knowledge of spondee vocabulary (N=96).

Threshold	*Knowledge of Spondee Vocabulary*	
	Without Prior Knowledge	*With Prior Knowledge*
1000 cps	8.6	9.3
Spondee	23.0	20.3

to measurement of the speech threshold, while the remaining subjects were not given this opportunity to familiarize themselves with the test items.

Each subject was seen in a single experimental session, during which two thresholds were measured, the threshold for a 1000-cps pure tone and the threshold for spondee words. The latter was obtained with recorded list E of CID Auditory Test W-1.

In order to visualize the effect of each factor separately, Tables 2 through 6 were prepared to present the mean 1000-cps and spondee thresholds obtained for the two categories of each factor in turn. Table 2, for example, illustrates only the effect produced on each threshold by varying the method of measurement. The mean spondee threshold for subjects in the clinical method group is only 0.1 db different from the mean threshold for subjects in the up-and-down method group. There is, however, a 1.7-db difference between the two methods for the 1000-cps thresholds. This difference is almost the exact order of magnitude to be expected in view of the difference in size of intensity steps used in the two methods. For the spondee threshold 2-db steps were used for both the clinical and up-and-down methods. For the 1000-cps threshold, however, 2-db steps were used in the up-and-down method, but 5-db steps were used in the clinical method. A theoretical difference of 1.5 db, in the direction of lower threshold intensity for the method involving 2-db steps would therefore be predicted. This difference derives from the fact that, when 5-db steps are used, the mean threshold is underestimated by 2.5 db, whereas, when 2-db steps are used, the mean threshold is underestimated by only 1.0 db. The observed difference of 1.7 db appears to be in relatively good agreement with this theoretical expectation.

Thus, for both the spondee threshold and the 1000-cps threshold, after allowance has been made for differences in the size of intensity steps used, there appears to be very little difference between results obtained by a clinical versus a laboratory procedure.

Table 3 shows that there is no large or systematic effect of sex on the auditory acuity of young normals. Females average 0.7 db better for the 1000-cps threshold, but males are 0.7 db better for the spondee threshold. Both differences are small.

Table 4 reveals a slight advantage (1.3 db) for the right ear at 1000 cps. However, this effect is not maintained for the spondee threshold, so that evidence of a slight tendency for one ear to have greater acuity is lacking.

Table 5 suggests that the effect of test order is not critical. The 1000-cps threshold SPL is slightly lower (1.5 db) when it is obtained first, but the spondee threshold SPL is slightly higher (1.3 db) when it is measured first.

Table 6 summarizes the effect of prior knowledge of the spondee vocabulary on the spondee threshold. The threshold SPL is 2.7 db lower for subjects who were read the complete list of 36 words prior to the test than for subjects not given prior knowledge. The possibility of a sampling error, such that subjects assigned to the prior knowledge group had inherently keener threshold sensitivity than subjects in the group without prior knowledge, is minimized by the fact that these same

TABLE 7. Summary of analysis of variance for 1000-cps threshold data.

Source		*df*	*ms*	*F*
Main Effects				
Method	(a)	1	63.7	3.56
Sex	(b)	1	11.1	0.62
Ear	(c)	1	42.7	2.38
Order	(d)	1	61.1	3.41
Prior Knowledge	(e)	1	8.9	0.50
Interactions				
c x e		1	255.5	14.26*
Within Sub-Classes		64	17.9	

*Significant beyond the 1% level. All other interactions were nonsignificant and are not reported here.

subjects did not yield a lower threshold SPL for 1000 cps. The mean 1000-cps threshold SPL for the prior knowledge group was, in fact, 0.7 db higher than the mean 1000-cps threshold SPL for subjects without prior knowledge of the spondee vocabulary.

This 2.7-db difference between spondee threshold SPL for subjects with and without prior knowledge of the spondee vocabulary was the largest effect observed over the five factors studied, and, as subsequent discussion shows, was the only effect that could be considered significant on the basis of statistical analysis.

Tables 7 and 8 summarize analyses of variance performed on the 1000-cps and spondee threshold data, respectively. Inspection of Table 7 shows that, for the 1000-cps threshold, none of the five factors studied appears to exert a significant effect, and there are no particular interactions worthy of note. Only one F ratio, that for the interaction between the ear tested and prior knowledge of spondee vocabulary (c x e), is significant beyond the 1% level (F=7.08). Since it is difficult to conceive a logical basis for an actual interdependence between these two factors to exist, insofar as the threshold for 1000 cps is concerned, the authors choose to conclude that this F ratio represents that chance occurrence under the null hypothesis which will, in fact, happen 1% of the time in a large series of F ratios computed on the basis of pairs of samples drawn from identical populations.

TABLE 8. Summary of analysis of variance for spondee threshold data.

Source	*df*	*ms*	*F*
Main Effects			
Method	1	0.0	0.00
Sex	1	13.3	1.21
Ear	1	0.9	0.08
Order	1	42.8	3.89
Prior Knowledge	1	169.3	15.41*
Interactions†			
Within Sub-Classes	64	11.0	

*Exceeds value required for the 1% level of confidence.
†None of the interactions are reported since none exceeds value required for the 1% level of confidence.

Table 8 shows a similar, relatively negative, set of results for the spondee threshold analysis. Again, only one F ratio is significant beyond the 1% level. Here however, it is a main effect, knowledge of test vocabulary, with a difference of 2.7 db between spondee thresholds for subjects with and without prior knowledge of test vocabulary, which is significant. This finding is interpreted as evidence that preliminary exposure of audiologically naïve subjects to a list of spondee words lowers the measured threshold SPL slightly (2 to 3 db). All other F ratios obtained in the analyses of variance for spondee thresholds were nonsignificant. This situation, coupled with the totally negative findings in the companion analysis of the data on acuity for 1000 cps leads to the conclusion that systematic

TABLE 9. Mean 1000-cps threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) for each method of measurement and mean spondee threshold sound pressure levels for each method of measurement and for subjects with and without prior knowledge of spondee test vocabulary.

Method	*1000-cps Threshold*	*Spondee Thresholds*		
		Without Prior Knowledge	*With Prior Knowledge*	*Combined*
Up-and-Down	8.1	22.3	20.9	21.6
Clinical	9.8	23.6	19.7	21.7
Combined	9.0	23.0	20.3	21.6

effects due to method of testing, sex, ear, and order of threshold measurement are either nonexistent or are so small that they are obscured by the uncontrollable variables in the investigation. In either event, the implication is clear. These four factors did not produce effects so large that the effects modified the measured threshhold levels substantially. Hence, the influence of sex, ear, and order of test may be disregarded in examining the data at hand with the aim of assessing the relationship between acuity for 1000 cps and acuity for spondees. Technically, the same conclusion applies to the effect of method of test, but common sense argues that the size of the interval used in testing (which differed for pure tones in the two methods of test) should not be ignored completely. The rationale underlying this last statement and the bases for other general conclusions are highlighted by Tables 9 and 10.

Table 9 gives mean thresholds, subdivided in terms of method of test and, for the spondee thresholds, the factor of prior knowledge. Table 10 reports the differences, for the same breakdown of data, between the 1000-cps and the spondee thresholds. Appropriate combined values are also reported in both tables. The following conclusions seem pertinent and reasonable.

First, as seen in Table 9, and as mentioned earlier, the mean thresholds for 1000 cps appear essentially equivalent when allowance is made for the fact that the clinical method used a 5-db step as contrasted to the 2-db step employed in the up-and-down method. Thus, it would appear that this threshold can serve as a stable point of reference when the size of the test interval is specified.

Second, the mean thresholds for spondaic words varied appreciably. As already pointed out, the variation with method of testing can be considered

TABLE 10. Difference between mean 1000-cps threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) and mean spondee threshold sound pressure levels for each method of measurement and for subjects with and without prior knowledge of test vocabulary.

Method	*Prior Knowledge*		*Combined*
	Without	*With*	
Up-and-Down	14.2	12.8	13.5
Clinical	13.8	9.9	11.9
Combined	14.0	11.3	12.6

random, but the variation due to familiarity with the spondees is systematic. The important point is the fact that the speech threshold is not a point of reference whose stability is comparable to that of the 1000-cps threshold. Therefore, the establishment of an audiometric norm for speech requires the designation of such additional conditions as: (a) the specific test material on which the norm is based and (b) the audiological sophistication of the subjects who are the reference group.

Third, as the foregoing conclusions imply and as Table 10 illustrates, the difference between the thresholds for the 1000-cps pure tone and for the W-1 spondees varies substantially. This variation is primarily the result of the instability of the spondee threshold, although the change in the size of the interval used to measure acuity for 1000 cps is thought to have exerted its influence as well. The outstanding point is that the observed differences range from 9.9 to 14.2 db, with the average for all conditions combined being 12.6 db. These differences are all substantially greater than the 6-db value currently designated in the American Standards for Speech Audiometers (*2*). They are in reasonable agreement with several earlier investigations (*3*, *4*, *6*), and smaller than reported by other writers (*9*, *11*).

Discussion

The practical implications of the present study have already been partially stated. These implications are: (a) that a multiplicity of conditions must be specified in order to stabilize the norm for speech audiometry, (b) that the difference between the norms for pure-tone audiometry and for speech is a function of the conditions chosen in specifying both but (c) that the difference between the norms for 1000 cps and for speech should be designated as substantially more than 6 db.

The magnitude of the difference which is selected as specifying the standard relation between the threshold for 1000-cps pure tone and threshold for speech must be settled by arbitrary decision of the persons responsible for establishment of standards. The present study can assist these persons only to the degree that it helps to clarify the factors to be considered.

To this end, it is instructive to examine the differences between thresholds exhibited by the 20 subjects used in the preliminary study (see Table 1) and to compare these results with the findings already reported for the 96 subjects who participated in the main experiment.

It probably represents approximately the limiting range to be encountered when subjects involved have 'normal' acuity. In other words, the evidence at hand leads one to the belief that, when the size of the test interval is constant for both measures, highly sophisticated listeners will detect a 1000-cps pure tone at a sound-pressure level about 10 db weaker than the level at which they correctly repeat 50% of the W-1 words. This difference may become as great as 17 db in consequence of complete unfamiliarity with the spondee test materials. Since the threshold for 1000 cps is but little affected by audiological sophistication, the difference may fluctuate over a range of 7 or 8

db as a result of variation in the threshold for speech.

Assuming that the foregoing analysis is correct, three choices are available in choosing a 'standard' difference to be incorporated in a revision of audiometric norms.

If highly sophisticated listeners are selected as the criterion population, a difference of about 10 db must be specified. The practical consequence of such a choice will be that the naïve listener (including many a hard-of-hearing person) will yield initial thresholds which are several decibels poorer than his later ones will be, particularly if time is not taken to familiarize him with the test words prior to the initial test.

If the fully naïve listener is chosen as the standard, the reverse situation will exist. A difference of 15 or 16 db must now be specified, and any person having appreciable prior experience with the speech material will obtain thresholds which appear better by several decibels than they otherwise would.

If a difference of intermediate value (12 or 13 db) is specified, this choice would imply that a moderately sophisticated listener is the criterion, and it would keep the discrepancy small (*circa* 3 db) in the measurement of threshold for other types of listeners.

This choice (that is, establishing the difference between the norm for 1000 cps and the norm for speech at 12 to 13 db) would seem to be most reasonable. It represents a 'middle-of-the-road' course if one is thinking in terms of spondee words as represented by W-1 recording and also if one is contemplating the array of evidence which studies other than the present one have supplied. Moreover, since the spondaic words have relatively high audibility, even audiologically sophisticated listeners will exhibit a difference between modalities of at least 12 or 13 db when the speech threshold is determined with other types of material.

Summary

In order to determine the intensity difference between normal hearing for spondee words and normal hearing for a 1000-cps pure tone, both types of threshold were measured in 10 audiologically sophisticated and 106 audiologically naïve listeners.

The threshold for 1000 cps was found to be relatively independent of the sophistication factor, but prior familiarity with the spondee test vocabulary exerted a significant influence on the spondee threshold.

Results indicate that the intensity difference between normal hearing for a 1000-cps pure tone and for spondee words is considerably larger than 6 db. It ranged from 9.4 to 17.1 db depending on variations in the type of subject, the familiarity of speech materials, and aspects of the measurement procedure. A difference of 13 db is approximately medial. It represents a value which might properly be selected as the relationship to be specified for audiometric standards.

References

1. *American Standard Sound Measurement Package, American Standard Specification for Audiometers for General Diagnostic Purposes,* appr. March 21, 1951. American Standards Association, Inc., 70 East 45th St., New York 17, N. Y.

2. *American Standard Sound Measurement Package, American Standard Specification for Speech Audiometers,* appr. July 10, 1953. American Standards Association, Inc., 70 East 45th St., New York 17, N. Y.
3. Corso, J. F., Confirmation of the normal threshold for speech on C.I.D. auditory test W-2, *J. acoust. Soc. Amer.,* 29, 1957, 368-370.
4. Corso, J. F., and Cohen, A., Methodological aspects of auditory threshold measurements. *J. exp. Psychol.,* 55, 1958, 8-12.
5. Dadson, R. S., and King, J. H., A determination of the normal threshold of hearing and its relation to the standardization of audiometers. *J. Laryngol. Otol.,* 66, 1952, 366-378.
6. Davis, H., The articulation area and the social adequacy index for hearing. *Laryngoscope,* 58, 1948, 761-778.
7. Davis, H., and Usher, J. R., What is zero hearing loss? *J. Speech Hearing Dis.,* 22, 1957, 662-690.
8. Dixon, W. J., and Massey, F. J., *Introduction To Statistical Analysis.* New York: McGraw-Hill, 1951.
9. Glorig, A., Wheeler, D., Quiggle, R., Grings, W., and Summerfield, A., *1954 Wisconsin State Fair Hearing Survey*—statistical treatment of clinical and audiometric data. American Academy of Ophthalmology and Otolarygology, 1957, Research Center, subcomm. Noise in Industry, Los Angeles.
10. Harris, J. D., Normal hearing and its relation to audiometry. *Laryngoscope,* 64, 1954, 928-957.
11. Lightfoot, C., Carhart, R., and Gaeth, J. H., Masking of impaired ears by noise. *J. Speech Hearing Dis.,* 21, 1956, 56-70.
12. National Health Survey, 1935-1936. Preliminary reports, hearing study series, bull. 5, *Normal Hearing for Speech at Each Decade of Life.* Div. Pub. Health Methods, Nat. Inst. Health, U. S. Pub. Health Serv., Washington, 1938.
13. Newby, H. A., *Audiology, Principles and Practice.* New York: Appleton-Century Crofts, 1958.
14. Newhart, H., and Reger, S. N., (Eds.), *Manual for a School Hearing Conservation Program* (rev. 1951, for Comm. on Conservation of Hearing, C. E. Kinney, G. D. Hoople, S. R. Guild, S. N. Reger, Eds., suppl., trans. Amer. Acad. Ophth., Otolaryng., 100 First Ave. Bldg., Rochester, Minn.). Omaha: Douglas Printing Co.

Some Factors Affecting the Spondee Threshold in Normal-Hearing Subjects

TOM W. TILLMAN

JAMES F. JERGER

The effects of subject sophistication on auditory test results vary with the particular stimulus being employed to assess auditory acuity. For example, Jerger and others (2) found that essentially equivalent pure-tone thresholds are obtained for highly sophisticated, highly practiced listeners and for listeners who have had no previous experience in pure-tone audiometric examinations. However, the same experiment suggested that the spondee threshold SPL for sophisticated subjects, who were relatively familiar with the spondee vocabulary, was some 5 to 7 db lower than that yielded by inexperienced subjects who were not familiar with the test vocabulary as such. It was found, moreover, that if spondee threshold measurements were repeated in these two groups, the second threshold SPL for the inexperienced subjects was approximately 3 db lower than the initial threshold SPL, while there was essentially no difference in the two thresholds for the sophisticated group.

The above findings led these same investigators (2) to formulate a more extensive experiment utilizing 96 unsophisticated normal-hearing subjects between the ages of 18 and 24 years. The design of the experiment allowed the assessment of the effect of prior knowledge of the test vocabulary on the spondee threshold. Prior knowledge of test vocabulary was imparted to half of the subjects by reading the entire list of spondee words to them immediately preceding the measurement of the spondee threshold. The mean spondee threshold SPL for the group given previous knowledge of the test vocabulary was approximately 3 db better than that for the group not given such knowledge.

Since the 3-db improvement observed in the preliminary experiment could have been due, in part, to knowledge of test vocabulary and, in part, to practice in the task of responding to words at threshold intensities, the present study was designed to isolate the effects of these two variables: (a) knowledge of spondee test vocabulary, and (b) practice in the task of responding to faint spondee words.

Tom W. Tillman (M.A., University of Oklahoma, 1955) is Research Assistant, Northwestern University. James F. Jerger (Ph.D., Northwestern University, 1954) is Assistant Professor of Audiology, Northwestern University. This study was supported by funds provided under Contract AF41(657)-185 with the USAF School of Aviation Medicine, Randolph Field, Texas.

Apparatus

The core of the apparatus used to measure spondee and 1000-cps thresholds consisted of a commercially available speech audiometer (Grason-Stadler, type 162) feeding a PDR-10 earphone mounted in an MX41/AR cushion. All spondee thresholds were obtained by playing portions of recorded List E of CID Auditory Test W-1 through this speech audiometer.

A separate pure-tone source, consisting of a commercially available clinical audiometer (Beltone, model 15-A), was fed through the speech audiometer to the same PDR-10 earphone. This portion of the apparatus was utilized in obtaining thresholds for a pure tone of 1000 cps.

The electrical signal across the earphone was measured periodically with a vacuum-tube voltmeter (Hewlett-Packard, model 400A). The acoustic output of the apparatus was calibrated by means of an ASA Type-1 coupler, calibrated condenser microphone (Western Electric, type 640AA), cathode follower (ADC, D5153), and vacuum-tube voltmeter (Hewlett-Packard, model 400A). All speech and pure-tone thresholds reported below are, therefore, expressed as the sound pressure level, reference 0.0002 microbar, developed in an ASA Type-1 coupler.

The sound pressure level developed by the earphone at octave frequencies from 125 to 8000 cps was measured before beginning the experimental testing and at three-day intervals throughout the course of the testing. The maximum variation in output over the testing period at the test frequency of 1000 cps was 0.5 db. This variation equalled or exceeded the maximum variation which occurred at any octave frequency from 250 to 4000 cps. However, maximum variations in output of 0.6 and 2.5 db occurred during the testing period at 125 and 8000 cps, respectively. Prior to the experiment, measurements demonstrated that the speech audiometer equalled or exceeded all specifications listed in the American Standard Specifications for Speech Audiometers (*1*).

Subjects and Procedure

Three groups of 10 subjects each were selected from the female student body at Northwestern University. The age range of each group was restricted to 18 to 24 years, and no subject was accepted who reported any history of ear pathology, excessive noise exposure or previous experience as a listener in auditory tests involving speech stimuli. All prospective subjects were screened audiometrically at a hearing level (hearing loss dial setting) of 10 db, reference USPHS norm, at each octave interval from 125 to 8000 cps, as well as at 1500, 3000, and 6000 cps. No subject was accepted who failed to respond to any test frequency at the screening intensity in either ear. Only one ear of each subject was used in the collection of the data reported below. The right ear was tested in half of the subjects, the left ear in the remainder. Subjects accepted for the experiment were assigned to the three experimental groups in a haphazard and nonsystematic fashion.

Each subject was seen in a single experimental session during which three thresholds were measured, the threshold for 1000-cps pure tone and two

separate thresholds for spondee words. The measurement of the threshold for the 1000-cps tone always followed the initial spondee threshold and preceded the second spondee threshold.

Group A (Practice Only). One group of 10 subjects, hereafter referred to as Group A, was included to test for the effects of practice only. In this group the first spondee threshold was obtained by playing words 1 through 18 of recorded List E of CID Auditory Test W-1. The second spondee threshold was obtained by playing words 19 through 36 of the same recording. The effects of practice alone on the spondee threshold could then be evaluated by comparing the mean threshold obtained in the first test with the mean threshold obtained in the second test.

Group B (Practice and Possible Prior Knowledge). A second group of 10 subjects, hereafter referred to as Group B, was included to assess the combined effects of practice and possible knowledge of test vocabulary, prior to the second threshold. In this group, the first spondee threshold was obtained by playing words 1 through 18 of recorded List E of CID Auditory Test W-1. The second spondee threshold was then obtained by playing the same 18 words again. The effects of practice plus possible knowledge of test vocabulary could then be evaluated by comparing the magnitude of the two mean thresholds thus obtained.

Group C (Practice and Definite Prior Knowledge). A third group of 10 subjects, Group C, was formed to evaluate the combined effects of practice and definite prior knowledge of test vocabulary on the spondee threshold. In this group, both spondee thresholds were obtained by playing words 1 through 18 of recorded List E of CID Auditory Test W-1. In addition, each subject was given prior knowledge of the test vocabulary by hearing the first 18 words of this list, in the manner described below, immediately preceding the measurement of the first spondee threshold. The effects of definite prior knowledge of test vocabulary could then be evaluated by comparing the mean initial threshold for this group with the mean initial thresholds obtained for Groups A and B.

All spondee thresholds were obtained by the up-and-down method, described in detail by Jerger and others (*2*), utilizing 2-db steps of attenuation. This method was chosen in preference to a simpler and less time-consuming clinical procedure because it allowed each threshold to be defined using exactly the same number of stimulus presentations. In other words, each spondee threshold was defined by presenting a series of 18 spondee words using the up-and-down method of threshold measurement. As noted earlier, these 18 words were always either words 1 through 18 or words 19 through 36 of recorded List E of CID Auditory Test W-1.

A clinical procedure utilizing 2-db steps of attenuation was employed to measure each subject's threshold for a 1000-cps pure tone. The procedure employed was the ascending technique described in the 1951 revision of the 'Manual for School Hearing Conservation Programs' (*3*) prepared by the Committee on Conservation of Hearing of the American Academy of Ophthalmology and Otolaryngology.

TABLE 1. Mean spondee and 1000-cps threshold sound pressure levels (db *re:* 0.0002 microbar in ASA type-1 coupler) and their standard errors for three groups of 10 normal hearing subjects each: Group A, practice alone; Group B, practice and possible prior knowledge; Group C, practice and definite prior knowledge.

	First Spondee Threshold	*Second Spondee Threshold*	*1000 cps*
Group A			
Mean Threshold SPL	25.9	24.8	9.3
Standard Error of Mean	1.0	1.3	1.3
Group B			
Mean Threshold SPL	27.0	24.6	10.9
Standard Error of Mean	1.2	1.1	1.6
Group C			
Mean Threshold SPL	20.9	20.6	10.7
Standard Error of Mean	1.0	0.6	1.2

The instructions read to subjects in the present study preceding the experimental session were essentially identical to those used by Jerger and others (*2*). The only changes were necessitated by the fact that the latter study employed only two test runs, while the present study used three test runs.

In addition to the instructions noted above, which were read to all subjects, the following instructions were read to the prior knowledge group immediately before the initial spondee threshold was measured.

> Before the initial word test, I will read a series of 18 two-syllable words at a level which is easy for you to hear. You are to repeat back each word. These words are the same ones that you will later hear in the two word tests; however, they will be in a different order. Since the purpose of this initial reading is to make you familiar with the words, please listen carefully.

The first 18 words of List E of CID Auditory Test W-1 were then pronounced in alphabetical order by monitored live voice at a level 50 db above the present American norm of 22 db SPL.

Results

Table 1 summarizes the experimental data. Group A, in which the first spondee threshold was obtained using the first 18 words of recorded List E and the second threshold using the second 18 words of the same recorded list, showed an improvement in threshold acuity of 1.1 db from the first to the second test. This is a very slight difference, and it would seem hazardous to conclude that it resulted from practice in the task of responding to spondee words at threshold intensities. This seems particularly true in view of the essentially equivalent thresholds for the two spondee threshold tests in Group C. In this latter group, the two thresholds were obtained using the first 18 words of recorded list E after the subjects had been given previous experience with the test words. The 0.3-db improvement in this group from the first to the second test could hardly indicate an important practice effect.

In Group B an improvement of 2.4 db occurred from the first to the second spondee threshold test. The two

thresholds in this group were obtained using the first 18 words of recorded List E. This group, it will be recalled, was included to assess the effect of possible knowledge of test vocabulary imparted to the subjects during the first spondee test as well as the effects of practice. In view of the relatively small change which occurred from the first to the second thresholds in Group A and the essentially equivalent thresholds obtained in Group C, it seems reasonable to conclude that the improvement noted in Group B was due to the knowledge of test vocabulary gained by the subjects as a result of the initial test.

Comparison of both the first and second spondee thresholds obtained in Group C (definite prior knowledge of the test vocabulary) with those for Group A (no prior knowledge) reveals that Group C yielded threshold SPLs 4 to 5 db lower than Group A. It seems reasonable to attribute this relatively large difference solely to the fact that Group C had been given prior knowledge of the test vocabulary.

The above findings suggest that, at least insofar as short-term effects are concerned, practice in the task of responding to spondee words at near-threshold intensities exerts no important influence on the spondee threshold in normal hearing subjects. Conversely, it would appear that previous knowledge of the spondee test vocabulary has an important effect. Specifically, threshold SPLs as much as 4 to 5 db lower will be obtained from subjects who are given previous knowledge of the test vocabulary than from subjects who are not given this prior knowledge.

These findings have practical implications for the clinician as well as for those concerned with establishing a normative standard for the hearing of speech stimuli. In the routine clinical procedures requiring successive measurements of the speech threshold using spondee words, the initial threshold SPL established may be 4 to 5 db higher than succeeding ones unless time is taken to familiarize the subject with the test vocabulary preceding the initial threshold test.

In establishing an audiometric norm for speech, where spondee words are specified as the stimulus, one must apparently be concerned with the extent to which the criterion population is familiar with the test vocabulary. Such concern is necessary because the absolute value of the standard, as well as the relationship between hearing for pure tones and for speech, will be influenced rather importantly by the degree to which the criterion population has prior familiarity with the test vocabulary.

Summary

The purpose of this experiment was to isolate the effects on the spondee threshold in normal-hearing subjects of (a) knowledge of spondee test vocabulary and (b) practice in the task of responding to spondee words at threshold intensities.

Two spondee thresholds were obtained for each of 30 subjects under varying experimental conditions utilizing recorded List E of CID Auditory Test W-1.

Results suggest two conclusions: (a) short-term practice in the task of responding to spondee words at thres-

hold intensities exerts no important influence on the spondee threshold SPL in normal-hearing subjects; and (b) normal-hearing subjects given prior knowledge of the test vocabulary yield spondee threshold SPLs which are 4 to 5 db lower than those yielded by subjects not given such knowledge.

References

1. *American Standard Sound Measurement Package, American Standard Specification for Audiometers for General Diagnostic Purposes,* approved March 21, 1951. American Standards Association, Inc., 70 East 45th St., New York 17, N. Y.
2. Jerger, J. F., Carhart, R., Tillman, T. W., and Peterson, J. L., Some relations between normal hearing for pure tones and for speech. *J. Speech Hearing Res.*, 2, 1959, 126-140.
3. Newhart, H., and Reger, S. N., (Eds.), *Manual for School Hearing Conservation Program* (rev. 1951, for Comm. on Conservation of Hearing, C. E. Kinney, G. D. Hoople, S. R. Guild, S. N. Reger, Eds., suppl., trans. Amer. Acad. Opth., Otolaryng., 100 First Ave. Bldg., Rochester). Omaha: Douglas Printing Co.

Validity of the One-Frequency Screening Principle for Public School Children

IRA M. VENTRY

HAYES A. NEWBY

The increasing acceptance afforded public school hearing conservation programs has pointed up the need for auditory screening devices that are efficient, economical, and practical. A recent development in screening tests has been the single-frequency test reported by House and Glorig (*2, 3, 4, 5*). The rationale for testing at one frequency (4000 cps) is that the loss at this frequency appears to be as great as or greater than the loss at any lower frequency. Davis, Hoople, and Parrack (*1*) have stated that an empirical rule about hearing loss seems to be that hearing is very seldom worse for any frequency important for speech than it is for 4000 cps. They also have said that the generality of this rule, *at least for adults* (present authors' italics), has only very recently been realized. Thus it would appear that testing at 4000 cps would detect hearing loss not only at this frequency but also at lower frequencies.

The purpose of this study was to investigate whether the generality of the single-frequency principle can be extended to include school children. Two null hypotheses were formulated for statistical testing of the principle involved: (a) there is no difference between the mean threshold losses from 500 through 8000 cps; (b) the mean threshold loss at 4000 cps is not different from the mean threshold losses above and below 4000 cps.

Procedure

With a standardized five-frequency sweep test, 1517 first, third, and fifth grade children were screened individually. The fail criterion for the sweep test was an inability to hear any one frequency in either ear at the screening levels of 20 db for 500 and 1000 cps and 15 db for 2000, 4000, and 8000 cps. Every fourth child failing the sweep test was given a pure-tone air-conduction threshold test at the above five frequencies on the ear(s) failing the sweep test. The threshold tests were administered by an experienced audiologist using a calibrated Maico E-2 audiometer. All threshold tests were conducted in the school rooms that had

Ira M. Ventry (Ph.D., Stanford University, 1958) is Research Audiologist, Speech Correction and Audiology Clinic, Veterans Administration Hospital, San Francisco. Hayes A. Newby (Ph.D., University of Iowa, 1947) is Professor of Speech Pathology and Audiology, Stanford University. This article is based on a portion of the Ph.D. dissertation written by Dr. Ventry and directed by Dr. Newby, and supported in part by a research grant from the National Institutes of Health, Institute of Neurological Diseases and Blindness.

been utilized for the sweep testing. Standard audiometric procedures were employed in the determination of thresholds. Special care, of course, was taken to insure the honesty and reliability of a subject's response.

A total of 62 subjects (90 ears) received threshold tests. The distribution of males and females was approximately equal with 32 males and 30 females. The mean age of all the children was 105.7 months with a range of from 77 to 139 months. Data were analyzed for each of the three grades separately and for all grades combined.

Results

Analysis of First Grade Threshold Data. Threshold tests were received by 23 first grade children, 14 males and nine females. Their mean age was 84.4 months with a range of from 77 to 93 months. A total of 31 ears received threshold tests and the analysis is based on these 31 ears[1]. McNemar's (*6*) Case VIII Mixed Model analysis of variance was employed to test the hypothesis (a) above. The mean losses at the five frequencies tested are shown in Table 1. Table 2 presents a summary of the analysis of variance, the results of which indicate that there are significant differences between the mean threshold losses from 500 through 8000 cps ($F=3.55$ and $P < .01$).

The second hypothesis (b) was then tested by evaluating the difference between the mean threshold loss at 4000 cps and the threshold loss at 1000 cps. Rejection of the hypothesis of no difference between 4000 cps and the next highest mean threshold loss (in this case 16.77 db at 1000 cps) indicates a significant difference between 4000 cps and each of the other frequencies. The difference between these two correlated means was tested by an F test[2]. The resultant F equaled 19.84 and indicated that the difference between the mean threshold loss at 1000 and at 4000 cps is highly significant ($P < .001$, $df=1$ and 30). The implications of these results are obvious and will be discussed later.

Analysis of Third Grade Threshold Data. Exactly the same analysis as the above was made on the threshold results of 22 third graders, 10 males and 12 females. These subjects had a mean age of 108.0 months with a range of from 101 to 122 months. Threshold tests were received by 34 ears. Table 1 presents the mean threshold losses at each of the five frequencies, and the results in Table 2 indicate that there is a highly significant difference ($P< .001$) between these mean losses. When the difference between 1000 and 4000 cps was subjected to statistical test, the resultant F of 39.44 was found to be highly significant ($P < .001$, $df = 1$ and 33).

[1]It might be argued that the use of two ears of one subject, at least in some cases, does not take into account the relationship between the ears of one subject. The same type of analysis, however, was employed utilizing only one ear of a subject (N subjects now equaled N ears). The statistical trend remained unchanged. For example, when the threshold results for all left ears were analyzed, the difference between 4000 cps and the frequency showing the next highest mean threshold loss was highly significant ($P<.001$). This same level of significance was reached when all the right ear threshold data were analyzed.

[2]The formula used in this case was $F = (\overline{X}.1 - \overline{X}.2)^2 / \Sigma(Dr - \overline{D})^2 / R(R - 1)$, where $n_1 = 1$ and $n_2 = R - 1$ (*6*).

TABLE 1. Mean threshold losses in db sensation levels for first, third, and fifth grade ears and for all ears combined.

Grade	*N*	*Frequencies (cps)* 500	1000	2000	4000	8000
1	31	16.29	16.77	14.35	21.77	15.97
3	34	22.06	22.21	20.59	31.18	17.79
5	25	15.40	14.00	14.00	23.60	11.20
All	90	18.22	18.06	16.61	25.83	15.33

Analysis of Fifth Grade Threshold Data. Threshold tests were received by 17 fifth grade subjects, eight males and nine females (25 ears). The mean age of these children was 131.4 months with a range of from 128 to 139 months. Table 1 presents the mean threshold losses from 500 through 8000 cps and Table 2 indicates that the difference between these mean losses is significant ($P < .01$). The F test comparing the mean threshold losses at 500 and 4000 cps resulted in a significant F of 6.47 ($P < .05$, $df = 1$ and 24).

Analysis of Combined Threshold Data. For the 62 first, third, and fifth grade subjects combined, Table 1 again illustrates the mean losses at each of the frequencies tested. Table 2 shows that the differences between the mean losses is highly significant ($P < .001$). The F test of the difference between the two greatest losses (500 and 4000 cps) resulted in an F of 47.04 which, once more, is highly significant ($P < .001$, $df = 1$ and 89).

Discussion

The results of the statistical analyses indicate that, for the population of school children from which the experimental group might be considered

TABLE 2. Analyses of variance for testing the differences between mean threshold losses for first, third, and fifth grade ears and for all ears combined.

Source	*df*	*ms*	*F*	*p*
First grade				
Frequency	4	245.21	3.55	.01
Ears	30	191.25		
Remainder	120	68.98		
Total	154			
Third grade				
Frequency	4	861.18	11.40	.001
Ears	33	800.59		
Remainder	132	75.56		
Total	169			
Fifth grade				
Frequency	4	553.20	4.53	.01
Ears	24	309.12		
Remainder	96	122.05		
Total	124			
All ears				
Frequency	4	1496.09	18.47	.001
Ears	89	487.87		
Remainder	356	81.02		
Total	449			

a random sample, the concept upon which one-frequency screening rests is a valid one. The two null hypotheses formulated—(a) that there is no difference between mean threshold losses and (b) that the loss at 4000 cps is not different from the loss at any other frequency tested—are rejected at statistically significant levels.

The validity of the one-frequency concept is given further support when another factor is analyzed. It will be noted (Table 1) that the frequencies showing the highest mean loss, excluding 4000 cps, are 500 and 1000 cps. It is commonly recognized, however, that hearing at these frequencies may be affected by ambient noise levels found in a testing environment, particularly in a school test situation. The same ambient noise levels have much less effect on hearing for the higher frequencies. The threshold tests conducted in this study took place, out of necessity, in an environment that was considerably noisier than the average test rooms employed in clinical audiometry. It is reasonable to assume that had the threshold tests been conducted in a quieter environment, even greater differences between the hearing losses at 500 and 1000 cps and the loss at 4000 cps would have been obtained. Yet, even in an unfavorable testing environment, the loss at 4000 cps remained significantly greater than the mean losses at these two lower frequencies.

One further comment might be made with respect to the previous analyses. The only frequencies included in this study were 500, 1000, 2000, 4000, and 8000 cps. An important research problem would be an evaluation of the generality of the one-frequency principle by the inclusion of 1500, 3000, and 6000 cps in the analysis of threshold results to determine whether the loss at 4000 cps is greater than the losses at these other frequencies as well. Evidence was obtained that the loss at 4000 cps is greater than the loss at at least four of the traditional octave test frequencies. Whether the inclusion of other test frequencies, particularly the higher ones, would call for a revision of the principle remains to be explored. The importance of the need for this research is indicated by the growing emphasis on 3000 and 6000 cps as standard audiometric test frequencies.

Summary

The single-frequency screening principle was evaluated for school children by conducting pure-tone air-conduction threshold tests on every fourth subject failing a standardized five-frequency sweep test. A total of 1517 first, third, and fifth grade children received individual sweep tests; threshold results were obtained on 90 ears (62 subjects). The results of the threshold analyses indicated that there were significant differences between the mean threshold losses from 500 through 8000 cps. Further statistical analyses provided evidence that the mean threshold loss at 4000 cps is greater than the mean threshold loss at any of the other four frequencies tested. The validity of the single-frequency principle, at least under the conditions of this study, appears to be established. One-frequency screening, however, needs to be further evaluated to determine the efficiency of this type of screening in detecting medically significant hearing loss in children.

References

1. Davis, H., Hoople, G., and Parrack, H. O., The medical principles of monitoring audiometry. *A.M.A. Arch. industr. Health*, 17, 1958, 1-20.
2. Glorig, A., and House, H. P., New, practical concept of hearing testing for use of the general physician. *J. Amer. med. Ass.*, 166, 1958, 1719-1721.
3. Glorig, A., and House, H. P., A new concept in auditory screening. *A.M.A. Arch. Otolaryng.*, 66, 1957, 228-232.
4. House, H. P., A new instrument and concept for the rapid detection of a possible hearing impairment. *Trans. Amer. Acad. Ophthal. Oto-laryng.*, 61, 1957, 228-230.
5. House, H. P., and Glorig, A., A new concept of auditory screening. *Laryngoscope*, 67, 1957, 661-668.
6. McNemar, Q., *Psychological Statistics* (rev. ed.). New York: J. Wiley, 1955.

Dysphasic Speech Responses to Visual Word Stimuli

GERALD M. SIEGEL

This study was designed to investigate the speech behavior of dysphasic subjects in terms of certain operationally defined aspects of words. These aspects, or word parameters, were so chosen that generalizations might be made concerning various theories of dysphasia.

Considerable research in dysphasia has indicated that there are characteristic patterns of difficulty in the speech of dysphasic patients. Investigators have attempted to analyze dysphasic speech performance in terms of aspects of language and to identify those language parameters that typically create difficulty for the patient. Where such characteristic patterns of dysphasic difficulty have been found, or postulated, attempts have been made to explain them in terms of a larger theoretical framework concerning the entire problem of dysphasia.

Gerald M. Siegel (Ph.D., University of Iowa, 1957) is Instructor of Speech and Director of the Speech Clinic, North Dakota Agricultural College. This article is based on a Ph.D. dissertation completed under the direction of Dr. Frederic L. Darley. Completion of the research was made possible by a grant from the Speech Correction Foundation. A portion of this material was presented at the 1958 Convention of the American Speech and Hearing Association, New York City.

Review of Literature

Recency of Learning. Freud (*9*, pp. 87-89), Nielsen (*17*, p. 18) and Weisenburg and McBride (*26*, p. 96) have described dysphasia as being characterized in part by specific impairment of relatively recently learned material in contrast to more remote learning. An outgrowth of this theory is the contention that bilingual dysphasics typically retain greater ability in their mother tongue than in a more recently acquired language.

Difficulty of the Task. It has been suggested also that a factor contributing to dysphasic speech performance is the relative difficulty of the speech task. Thurston (*23*) found that on an oral spelling task dysphasic subjects had a disproportionate increase in number of errors as the words became more difficult.

Part of Speech. Several writers have indicated that dysphasics have particular difficulty in handling certain parts of speech. Freud (*9*, p. 34), Goldstein (*11*, p. 246), Weisenburg and McBride (*26*, p. 299), and Nielsen (*17*, p. 71) all describe a condition of 'amnesic aphasia' which is characterized by disturbance of the ability to use nouns and, in more severe cases, other parts of speech. Wepman (27, p. 236) describes a condition of 'anomia' which

involves difficulty in the use of common names. In the same source Wepman (*27*, p. 153) describes the speech recovery process as one in which common names for things (nouns) tend to be the first relearned patterns.

Eisenson (*8*, p. 14) attributes the supposed increased difficulty with nouns to the high frequency of occurrence of this particular part of speech in normal language rather than to any peculiarity of the nouns themselves. Brown and Schuell (*3*) and Schuell (*21*) have argued that dysphasics suffer generalized impairment of the ability voluntarily to select any specific word, regardless of the part of speech. Similarly, Olhoeft's (*18*) dysphasic patients failed to demonstrate, in ability to utter words, substantial differences among five parts of speech.

Wepman and others (*28*) recently compared a sample of the verbal output of one 'anomic' patient with that of a normal subject and a derived normal model of spoken English. They reported that for their 'anomic' patient there was a tendency for the retention of frequently occurring words and for relatively greater loss of less frequently occurring words when the verbal output of their dysphasic patient was compared to that of their control subject and the normal model. The authors suggest that the dysphasic subject's language repertoire may be related to a factor of 'word generality,' which is in turn related to 'frequency of occurrence.'

Specific Phonetic Factors. According to Head (*14*, p. 121) neither the part of speech nor the specific phonetic composition should be considered an independent factor in dysphasia. Several writers have, however, indicated that certain phonetic combinations may be more difficult than others for dysphasics. Goldstein (*11*, p. 43) and Huber (*15*) allude to specific phonetic difficulty while Anderson (*1*) and Schuell (*20*) have tended to discount the notion of any specific constellation of articulation errors among dysphasics. Schuell found articulation errors to be inconsistent and variable among dysphasics.

Levels of Word Abstraction. According to Goldstein (*11*, *12*) one of the sequelae of the brain pathology that precipitates dysphasia is impaired capacity for abstraction and the tendency to respond in a concrete manner and to concrete aspects of complex situations. This 'behavioral dedifferentiation,' as Goldstein calls it, is considered a total personality change involving marked reorganization of the patient's responses to abstract situations. Just as situations vary in inherent abstractness, words may conceivably vary with reference to level of abstraction. As words become increasingly abstract, they may become correspondingly more difficult for the dysphasic patient whose ability to handle abstract material is impaired.

Eisenson (7), Wepman (*27*), and Brown (*2*) have not concurred with Goldstein as to the prevalence of impairment of ability to abstract among dysphasics.

Evaluating the role of 'word abstraction level' in dysphasic speech is complicated by the fact that there is no generally accepted operational definition of this construct and, to the knowledge of the writer, no scale of level of abstraction of words.

Word Length. Attempts have been made to relate dysphasic speech errors to word length. Word length, however, is a complex variable that may be related to any of a number of either auditory or visual criteria or even to so remote a factor as grammatical part of speech. Writers have variously referred to the 'small words' of speech in terms of the number of letters and syllables or as connoting the so-called 'lesser' parts of speech (such as articles, conjunctions, and prepositions).

Goldstein (*11*, p. 70) and Wepman (*27*, p. 153), apparently referring to these lesser parts of speech, report that dysphasics have particular difficulty with 'small words.' Schuell (*20*) also found that there may be increased difficulty with small words but she related this to the auditory or visual vividness and prominence of words rather than to part of speech.

It is an interesting fact that for both Schuell and Goldstein this factor of word length has meaning for a general theory of dysphasia. Goldstein explains it in terms of his theory of impairment of abstraction. He indicated that these so-called small words do not normally present difficulty for dysphasics when encountered in sentences since in this context they are concretely attached to either preceding or following words. When encountered in isolation, however, without contextual attachments, these words become quite abstract and therefore much more difficult for the dysphasic to handle.

Schuell, on the other hand, considers difficulty with small words in terms of her basic theory of the influence of impairment of 'reauditorization' on dysphasic speech. She suggests that short words are more difficult for dysphasics than longer words because short words tend to be indistinct in connected speech and are consequently more difficult to reauditorize.

The Problem. Various authors have attempted to present a lawful explanation for the patterning of speech errors in dysphasic verbalization. Dysphasic speech behavior has been related to the factors of recency of learning, relative difficulty of the task, part of speech, phonetic composition, level of abstraction, and word length.

The present investigation was designed to study the relationship between dysphasic speech behavior and three of these factors: part of speech, word length, and level of word abstraction. Specifically, the following question was posed: Do dysphasic individuals show consistent patterns of speech error in response to word stimuli which are classified according to (a) part of speech, (b) level of abstraction, and (c) word length?

Procedure

Subjects. The subjects were 31 patients diagnosed as dysphasic and were drawn from 11 hospitals and treatment centers. They ranged in age from 20 to 66 years with a median age of 51.3 years. Five subjects had a history of cortical trauma and 26 subjects had suffered a cerebro-vascular accident. Median time in school was 9.1 years with all but two subjects having completed eight or more years. The group consisted of 24 males and seven females all of whom correctly produced no fewer than 10% and no more than 90% of the 90 responses required.

Subjects were not screened for dysarthria because it was felt that there is no satisfactory technique for differentiating a so-called dysarthric response from a dysphasic response. Similarly, such categories as 'receptive,' 'expressive,' 'nominal,' and 'amnesic' were not considered since, as Weisenberg and McBride (*26*, p. 47) have pointed out, 'Nowhere in the aphasia literature is confusion more evident than in the classification of types of disorder.' The reliability and stability of even such widely used classification systems as those of Eisenson (*6*) and Wepman (27) have been questioned (*19*).

Stimulus Material. The stimulus material for this study consisted of a set of 90 words typed in uppercase letters one-fourth inch high on individual 4-by-6-inch white cards. The words were drawn from graded word lists (*4*, *10*) and were designated as being no higher than fifth grade level. By controlling educational achievement of the subjects and grade level of the words some control was exercised also over the difficulty and recency of learning of the stimuli.

Each of the 90 words was classified according to part of speech, word length, and level of abstraction. Within each parameter, subclassifications were defined so that there were three parts of speech (nouns, verbs, adjectives), two word lengths (short, long), and three levels of abstraction (low, medium, high).

Part of speech was determined by reference to a standard dictionary (*25*). Word length was established by a criterion involving number of letters and number of syllables. 'Long' words contained two or more syllables and six or more letters. 'Short' words were single-syllable words of four or fewer letters. Since level of abstraction is a construct for which there are no objective criteria, an operational procedure which is essentially an equal-appearing-interval scaling method (*24*) was employed. This procedure is described by Darley, Sherman, and Siegel (*5*).

The 90 words comprising the word list were chosen by selecting 30 words at each defined level of abstraction in such a way that for each abstraction level there was equal representation in terms of the other parameters: part of speech and word length. The word list was so devised that when the subclassifications of any one parameter were considered, the influence of the other two parameters was counterbalanced. Thus, there were 30 nouns, 30 verbs, and 30 adjectives; there were 45 short and 45 long words; there were 30 words at each of the three levels of abstraction.

Responses. The subjects were required to utter each of the 90 stimulus words as the cards upon which they were typed were presented in a random order. Subjects were tested individually at the various participating hospitals and centers. All testing sessions were conducted solely by the experimenter and were tape recorded. The subjects' responses were later replayed and evaluated by the investigator as being either right or wrong. If response latency was greater than 15 seconds or if no response at all was elicited, the response was scored incorrect. Beyond this, responses were scored wrong if they were judged by the investigator to be

TABLE 1. Sum and mean number of errors for each subclassification of part of speech, word length, and word abstraction level for 31 dysphasic individuals.

Classification	*Sum of Errors*	*Mean of Errors*
Part of Speech		
Nouns	323	10.42
Verbs	404	13.03
Adjectives	405	13.06
Length		
Short	478	15.42
Long	654	21.09
Abstraction Level		
Low	379	12.23
Medium	356	11.48
High	397	12.81

significantly deviant from the acceptable utterance of the words.

One week after the original judging the experimenter rescored the tape-recorded responses of 10 of the subjects (900 discrete responses) and correlated these ratings with the original judgments for these 10 subjects. The formula for the correlation of a four-fold point surface (*13*) was applied and a phi coefficient of .90 was obtained.

Results

Statistical Analysis. Obtained differences for each of the three parameters studied (level of abstraction, word length, part of speech) were evaluated by separate statistical analyses. For each analysis the criterion measure used was the number of wrong responses, and the level of significance selected was 5%. The balanced construction of the word list made possible the use of a treatments-by-subjects design (*16*) or, in the case of only two treatments, as in the word length analysis, the simple *t* test for related samples.

In Table 1 are presented the sum and the mean number of errors in each of the subclassifications for each parameter studied.

Parts of Speech. Data reported in Table 1 indicate that of the three parts of speech, adjectives incurred the greatest number of errors while nouns incurred the smallest number of errors. Subjects responded almost identically to verbs and to adjectives. The analysis of variance results, summarized in Table 2, indicate significant differences among parts of speech. The difference necessary for significance at the 5% level is 1.37. Error scores for both verbs and adjectives are significantly greater than for nouns. The difference between verbs and adjectives is not significant. These findings do not support the hypothesis that the nominal aspects of speech offer selective difficulty for dysphasics.

TABLE 2. Analysis of variance of the errors made by dysphasic subjects on nouns, verbs, and adjectives.

Source	*df*	*ms*	*F*
Part of Speech (P)	2	71.43	9.46*
Subjects (S)	30	142.58	
PS	60	7.55	
Total	92		

*F_{01} (df = 2 and 30) = 8.77.

TABLE 3. Analysis of variance of the errors made by dysphasic subjects on words of low, medium, and high levels of abstraction.

Source	*df*	*ms*	*F*
Abstraction (A)	2	13.63	7.25*
Subjects (S)	30	142.58	
AS	60	1.88	
Total	92		

*$F_{.05}$ (df = 2 and 30) = 6.35.

Length. The mean number of errors for the short words was 15.42 compared to 21.09 for long words (see Table 1). The obtained *t* of 4.35 is

highly significant ($t_{.01}$=3.65, df=30). It is evident that subjects had considerably more difficulty with long words than with shorter ones.

Level of Abstraction. The analysis of variance results summarized in Table 3 indicate significant differences among levels of abstraction. The difference necessary for significance at the 5% level is .72. The error scores (see Table 1) for both high-abstraction-level words and low-abstraction-level words are significantly greater than for medium-abstraction-level words. The difference between high and low levels of abstraction is not significant.

Discussion

In evaluating the results of this study it was clearly recognized that the particular investigative approach used is only one of the several that might have been employed and that in the very process of attempting rigorous definition of the stimuli and the subjects a certain amount of information was either ignored or lost and a certain amount of artificiality was introduced. The investigator was particularly aware of the fact that words in isolation do not necessarily affect dysphasic speech in the same way that words in context do.

Part of Speech. The results obtained are in conflict with the traditional notion that dysphasics suffer a selective impairment of the ability to use nouns, that is, some sort of 'anomic' or 'nominal' impairment. These results tend to corroborate studies (*3*, *8*, *18*, *20*, *21*) that dispute the hypothesis of specialized difficulty with one or another part of speech. The writer can offer no explanation, on the other hand, as to why nouns should have incurred fewer errors than did the other parts of speech. It seems no more logical to suggest a special facility with nouns than to suggest some special difficulty with nouns.

Word Length. There were significantly more errors on long than on short words. According to Goldstein (*11*), presentation of 'small words' in isolation tends to invoke more error responses from dysphasics than presentation of the same words in context because the lack of context increases the degree of abstraction of these words. In the present study, however, the abstraction levels of the two groups of words had been counterbalanced. If the salient feature of these small words is their increased abstraction, the process of counterbalancing abstraction level should have equated them in terms of their difficulty for dysphasics. The predominance of errors on long words indicates that this was not the case. It is true, of course, that the small words used were not the so-called 'lesser' parts of speech specified by Goldstein and that the procedures employed for specifying abstraction level do not necessarily relate to his concept of word abstraction.

Schuell (*20*) suggests that small words may be selectively difficult in contextual speech because they tend to be slighted and are given scant auditory prominence. In the present study, words were presented in isolation rather than in context. Schuell also has suggested that with increasing length of stimulus unit more dysphasic errors occur because of increased difficulty in reauditorizing. If this is so, one would expect that the longer words

would be more difficult to reauditorize. Such an expectation is borne out in the present study.

Level of Abstraction. The finding that high-abstraction-level words were more difficult for dysphasics than medium-abstraction-level words tends to support the Goldstein (*11*, *12*) hypothesis of impairment relative to degree of abstraction. Two reversals were found, however. High- and low-abstraction-level words did not differ significantly and low-abstraction-level words were missed significantly more often than medium-level words.

It was felt that these contradictory findings (and in some measure the findings of the part-of-speech analysis) might be related to some aspect of the stimulus words that had not been considered in the initial formulation of the word list. One such possibility was suggested in an article by Wepman and others (*28*). Wepman reported that his dysphasic patient had a marked tendency toward the excessive use of words which occur frequently in written English usage and the under use of those which occur less frequently. The word list in the present study was selected without reference to frequency of occurrence in language usage. In order to investigate this factor the frequency of occurrence of each of the 90 words in the list was ascertained from the Thorndike and Lorge (*22*) word list. Forty-eight of the stimulus words were found to be within the 2000 most frequently occurring words in written English usage. These were designated as comprising a 'frequent' level. The remaining 42 words were distributed among the third through the tenth thousand most frequently occurring words in written English usage and comprised an 'infrequent' level.

TABLE 4. The number of frequent (N=48) and infrequent (N=42) words falling into each subclassification of part of speech, length, and level of abstraction.

Subclassifications	*Frequent Words*	*Infrequent Words*
Part of Speech		
Nouns	19	11
Verbs	19	11
Adjectives	10	20
Length		
Short	28	17
Long	20	25
Abstraction Level		
Low	4	26
Medium	20	10
High	24	6

Since the number of words in each category differed, the means could not be compared directly. For each subject it was necessary to convert the number of errors made on frequent words to a proportion of the total number of frequent words in the list (48 words) and to convert the number of errors on infrequent words to a similar proportion using 42 as the basis for the proportion. These proportions were then treated as scores and a t test for related samples was applied. The value of t, 4.87, was highly significant ($t_{.01}=3.65$, $df=30$). As suggested by Wepman, there was greater impairment for words occurring less frequently in written language usage.

Since frequency of occurrence had not been counterbalanced in the original word list, interpretation of the data becomes complicated. The apportionment of the two frequency levels in each of the other subclassifications is presented in Table 4. If frequency of occurrence is a basic factor in determining dysphasic performance, it is apparent that this influence was unequally distributed.

Reviewing the part-of-speech analysis in the light of data for frequency of occurrence, it becomes evident that frequency of occurrence alone cannot account for the errors made on the three parts of speech. The subjects made more errors on verbs than nouns even though there were identical numbers of frequent and infrequent words within these two parts of speech. Further, despite the greater number of frequent words among verbs in contrast to adjectives, the number of errors committed on these two parts of speech was not significantly different.

It is typical of the English language that long words tend to be used less frequently than short ones (*29*). This fact is reflected in the word list used in the present study (Table 4). The relatively high incidence of errors made on both long and infrequent words may be the result of an interrelationship between these two factors. It would be interesting to evaluate independently the influence on dysphasic speech of length and frequency of occurrence.

The high and medium abstraction levels were approximately the same in terms of the distribution between the two frequency levels (Table 4); the effect of frequency may thus be considered counterbalanced in this comparison. The finding that medium-abstraction-level words were more difficult perhaps reflects a 'true' abstraction function in which the higher the level the greater the difficulty of the words.

The fact that there were 26 infrequent words in the low abstraction level in contrast to only six such words in the high abstraction level may indicate that the influence of the abstraction level factor in this comparison was counteracted, thus accounting for the lack of a significant difference in errors on high- and low-abstraction-level words.

In comparing words of low abstraction level with those of medium abstraction level, there were many more infrequent words in the low-abstraction-level group and it is again conceivable that the factor of frequency was responsible for the greater number of errors on low-abstraction-level words than on medium-abstraction-level words.

It may be noted that frequently used words tended to be judged as high in abstraction level. This is understandable since frequently occurring words must serve a number of language functions and, as such, have numerous referents. This characteristic is one of the basic criteria for designating a word as being high in abstraction level.

It appears that frequency of occurrence is related to, but not synonymous with, the other parameters defined in this investigation. Perhaps somewhere in this pattern of interrelationships resides a comprehensive principle of dysphasic language performance.

The influence of frequency of occurrence on the other parameters cannot be interpreted in this study; nor can the difference in errors obtained for the two frequency levels confidently be evaluated because of possible contamination by the original three parameters which were not counterbalanced with frequency. It does appear, however, that frequency of occurrence should be defined as a variable in studying dysphasic speech performance and that this factor deserves further investigation.

Summary

Speech behavior of 31 dysphasic subjects was investigated in terms of several word parameters: part of speech, level of abstraction, length, and frequency of occurrence in written English usage. The obtained results indicate that dysphasics make more errors on adjectives than on either verbs or nouns (no evidence of 'anomia' was found); more on long words than on short; more on words of both high and low abstraction level than on those of medium level; and more on words occurring infrequently in the language than on those occurring frequently.

References

1. Anderson, Jeanette O., Aphasia from the viewpoint of a speech pathologist. *J. Speech Dis.*, 9, 1944, 209-226.
2. Brown, I., Abstract and concrete behavior of dysphasic patients and normal subjects. *J. Speech Hearing Dis.*, 20, 1955, 35-42.
3. Brown, J. R., and Schuell, Hildred, A preliminary report of a diagnostic test for aphasia. *J. Speech Hearing Dis.*, 15, 1950, 21-28.
4. Buckingham, B. R., and Dolch, E. W., *A Combined Word List*. Boston: Ginn, 1936.
5. Darley, F. L., Sherman, Dorothy, and Siegel, G. M., Scaling of abstraction level of single words. *J. Speech Hearing Res.*, 2, 1959, 161-167.
6. Eisenson, J., *Examining for Aphasia*. New York: Phychological Corp., 1946.
7. Eisenson, J., Aphasics: observations and tentative conclusions. *J. Speech Dis.*, 12, 1947, 290-292.
8. Eisenson, J., *Examining for Aphasia* (rev. ed.). New York: Psychological Corp., 1954.
9. Freud, S., *On Aphasia*. New York: Int. Univ. Press, 1953.
10. Gates, A. I., *A Reading Vocabulary for the Primary Grades* (rev. ed.). New York: Teach. Coll., Columbia Univ., 1935.
11. Goldstein, K., *Language and Language Disturbances*. New York: Grune and Stratton, 1948.
12. Goldstein, K., and Scheerer, M., Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).
13. Guilford, J. P., *Psychometric Methods*. New York: McGraw-Hill, 1936.
14. Head, H., *Aphasia and Kindred Disorders of Speech*. New York: Macmillan, 1926.
15. Huber, Mary, A phonetic approach to the problem of perception in a case of Wernicke's aphasia. *J. Speech Dis.*, 9, 1944, 227-237.
16. Lindquist, E. L., *Design and Analysis of Experiments*. Boston: Houghton Mifflin, 1953.
17. Nielsen, J. M., *Agnosia, Apraxia, Aphasia* (2nd ed.). New York: Paul B. Hoeber, 1946.
18. Olhoeft, Joyce M., An investigation of the relationship between expressive aphasics' difficulties with different parts of speech. M.A. thesis, Univ. of Maryland, 1953.
19. Schotland, Henrietta, An evaluative study of two tests for aphasia. M.A. thesis, Univ. of Iowa, 1953.
20. Schuell, Hildred, Auditory impairment in aphasia: Significance and retraining techniques. *J. Speech Hearing Dis.*, 18, 1953, 14-21.
21. Schuell, Hildred, *The Minnesota Test for Differential Diagnosis of Aphasia* (res. ed.). Minneapolis: Univ. Print. Dept., Univ. of Minnesota, 1955.
22. Thorndike, E. L., and Lorge, I., *The Teacher's Word Book of 30,000 Words*. New York: Teach. Coll., Columbia Univ., 1944.
23. Thurston, J. R., An empirical investigation of the loss of spelling ability in dysphasics. *J. Speech Hearing Dis.*, 19, 1954, 344-349.
24. Thurstone, L. L., and Chave, E. J., *The Measurement of Attitude*. Chicago: Univ. of Chicago Press, 1929.
25. *Webster's Collegiate Dictionary* (3rd ed.). Springfield, Mass.: G. and C. Merriam, 1919.
26. Weisenburg, T. H., and McBride, Katharine E., *Aphasia: A Clinical and Psychological Study*. New York: Commonwealth Fund, 1935.
27. Wepman, J. M., *Recovery from Aphasia*. New York: Ronald Press, 1951.
28. Wepman, J. M., Bock, R. D., Jones, L. V., and Van Pelt, Doris, Psycholinguistic study of aphasia: a revision of the concept of anomia. *J. Speech Hearing Dis.*, 21, 1956, 468-477.
29. Zipf, G. K., *The Psycho-Biology of Language*. Boston: Houghton Mifflin, 1935.

Scaling of Abstraction Level of Single Words

FREDERIC L. DARLEY

DOROTHY SHERMAN

GERALD M. SIEGEL

Development of a quantitative and operational definition of 'word abstraction level,' one aspect of Siegel's (*14*) investigation of dysphasic speech performance, is the subject of this report. The Siegel investigation posed the question as to whether certain characteristic patterns of speech difficulty emerge when dysphasic individuals are required to respond to visual word stimuli which have been classified according to a number of parameters. One of the parameters selected for study was the level of abstraction of word stimuli.

Impairment of the 'abstract attitude' has been considered by some to be a primary consequence of brain damage and an essential aspect of the dysphasic syndrome. For Goldstein (*9*) the notion of impairment of the abstract attitude is an organizing concept which explains much of the dysphasic subject's behavior. 'In the concrete attitude we are given over passively and bound to the immediate experience of unique objects or situations. Our thinking and acting are determined by the immediate claims made by the particular aspect of the object or situation . . .,' while in the abstract attitude 'we transcend the immediately given specific aspect of sense impressions; we detach ourselves from the latter and consider the situation from a conceptual point of view and react accordingly' (*9*, p. 6). Goldstein believes that one of the sequelae of the brain pathology that precipitates dysphasia is this impaired capacity for abstraction:

> Organic pathology in patients with brain disease disintegrates human behavior in such a way that the capacity for abstract behavior is impaired to greater or lesser degree in the patient. Thus, he becomes more or less reduced to a level of concreteness . . . so that he can perform only those tasks which can be fulfilled in a concrete manner (*10*, p. 9).

The behavioral disintegration (or 'dedifferentiation,' as Goldstein calls it) that results from this impairment of abstraction is considered a total personality change in which the patient's responses to complex situations undergo marked reorganization, with the patient more typically functioning in the 'con-

Frederic L. Darley (Ph.D., University of Iowa, 1950) is Associate Professor of Speech Pathology and Audiology, University of Iowa. Dorothy Sherman (Ph.D., University of Iowa, 1951) is Associate Professor of Speech Pathology and Audiology, University of Iowa. Gerald M. Siegel (Ph.D., University of Iowa, 1957) is Instructor of Speech and Director of the Speech Clinic, North Dakota Agricultural College.

crete mode' described. Goldstein indicates that there are various degrees of both abstract and concrete behavior and that

> . . . it must be emphasized that the process of dedifferentiation does not entirely destroy the ability to generate ideas and thought. What is affected and modified is the way in which thoughts and ideas are dealt with. Thoughts do arise, but can become effective only in a concrete way (*9*, p. 7).

Goldstein's views have been accepted and supported by some writers (*1*, *2*, *16*), while others have questioned his analysis on methodological grounds (*11*, *12*, *18*) or because of contrary clinical or research observations (*3*, *5*, *12*, *17*).

The degree of a subject's capacity for abstraction may be reflected in given speech performances where degrees of abstraction are involved. It would seem a reasonable extension of the Goldstein hypothesis to expect dysphasics to respond selectively to words as a function of their level of abstraction, if 'level of abstraction' could somehow be acceptably defined. In the syndrome Goldstein describes as 'amnesic aphasia' the outstanding symptom reported is the inability to perform 'true' naming (in contrast to 'pseudo-naming,' which involves merely the association of a word with an object). This inability is expressed primarily in the use of nouns. According to Goldstein, the process of naming is quite an abstract one. Thus the difficulty which amnesic aphasics have in handling nouns may be thought to be directly related to impairment of the ability to abstract.

Eisenson (*6*, p. 2) has commented on dysphasic performance relative to word abstraction level:

> Aphasic difficulty is likely to increase in the realm of abstract words for which there is no basic objectivity. We cannot easily concretize such conceptual terms as *beauty*, *humorous*, *tragic*, *incongruous*, *liberal*, *hope*, *love*, *atomic energy*, etc. To explain these terms we would have to use other terms. Similarly, we would have great difficulty indicating objectively the meaning of such words as *if*, *however*, *about*. Most of us who develop language facility learn how to use these terms in conventional and appropriate ways. We reveal our meanings through them and understand other persons' meanings as expressed with these nonobjective terms. These, interestingly enough, are the very terms which cause the greatest difficulty, both in comprehension and production (spoken or written) for most aphasic patients.

Evaluation of the role of 'word abstraction level' in the speech of the dysphasic is complicated by the fact that there is no generally accepted operational definition of 'word abstraction level.' One may agree with Eisenson that the words cited are 'abstract,' but he provides no cues as to their relative level of abstraction nor as to how dysphasics respond to shades of abstraction. Typically, where the problem of abstraction has been considered, it has been related to performance skills other than speech.

Flesch (7, p. 384) suggested a quantitative means of specifying the abstraction level of a verbal passage. In considering this problem he wrote:

> Though the significance of the level of abstraction is widely recognized, no studies have been reported that attempt to estimate or measure it quantitatively. Discussions are usually limited to generalized distinctions between abstract and concrete words, illustrative examples, and such figurative devices as the semanticists' 'abstraction ladder.'

Continuing, Flesch assumes that 'level of abstractness can be ascertained by computing the ratio of certain parts

of speech to certain other parts of speech in written expression.' He arbitrarily designated certain parts of speech as 'definite' (concrete) and others as 'non-definite.' By computing the ratio of 'definite' to 'non-definite' words in a given passage, he obtained what he called his 'abstraction index.' Unfortunately no adequate rationale was provided for the designating of given words as either 'definite' or 'non-definite.' Though this procedure yields a quantitative value, there is no assurance that the criteria upon which it is based are valid. A more fruitful approach would seem to be first to define some operation by which the level of abstraction of individual words can be specified.

Johnson (*13*) has written extensively concerning the process of abstraction and has devised an 'abstraction ladder' of the sort Flesch possibly had reference to. Commenting on the diagram (*13*, p. 135) he presents, Johnson (*13*, p. 140) writes:

> It is to be clearly understood, of course, that the precise level of abstraction on which a statement—or a word—rests cannot be designated. Our diagram of the process of abstracting is not to be read like a thermometer. This is not to say that the notion of level of abstraction might not lend itself to investigation by means of fairly reliable rating scales. Such an investigation has not yet been undertaken, but it doubtless will be in due time.

So far as is known to the writers, no such scale had been formulated until the present investigation was undertaken.

Procedure

The method developed to provide a quantitative and operational definition of 'word abstraction level' consisted essentially of an equal-appearing-interval scaling procedure, after the method devised by Thurstone and Chave (*15*). A group of 35 judges, graduate students registered for courses in speech pathology or semantics, scaled the abstraction level of each of 572 words. The word list was comprised of 201 nouns, 201 verbs, and 170 adjectives; words that are frequently used as more than one part of speech were excluded. All words were drawn from the *Reading Vocabulary for the Primary Grades* by Gates (*8*) and *A Combined Word List* by Buckingham and Dolch (*4*); the words selected were graded as no higher than fifth grade by at least one of these sources.

Words were classified into two categories with regard to length, 'short' and 'long.' Although the specification of what is 'short' or 'long' is necessarily arbitrary, a criterion involving both number of letters and number of syllables was selected as most adequate; in the case of visual presentation these are probably the most prominent aspects of length for the reader; furthermore, these aspects can be precisely specified. 'Short' words, of which there were 267, were defined as single-syllable words of four or fewer letters. 'Long' words, of which there were 305, were defined as words of two or more syllables containing six or more letters.

Each of the judges received a packet of mimeographed materials containing a set of instructions and the 572 words arranged in random order on 8″ x 10″ sheets. The instructions contained specific directions regarding precedures

to be followed in making the ratings and a general discussion of the concept of word abstraction, which is here presented verbatim:

> In the present investigation we are attempting to determine the *level of abstraction* of a number of words. Specific words, whether nouns, verbs, or adjectives, vary in their level of abstraction. It will be your task to scale these words on the basis of your judgment of the level of abstraction of each of them.
>
> The level of abstraction of a word may depend on a number of considerations. Words tend to become more abstract as they become more general in their reference or as they tend to leave out details. Thus, when comparing the abstraction level of two such words as 'cow' and 'animal,' we might make our judgment on the basis of which of these two words contains the least specific information and tends to be a more general or a more inclusive category. A similar judgment may be made for verbs, such as 'walk' and 'travel.'
>
> Words also become increasingly abstract as they become further removed from a specific reference. In this respect those words are *least* abstract that can be defined only by example, by demonstrating what they are used to name or to stand for. The more abstract the word, the more difficult it is to find some means of 'pointing at' the meaning of the word.
>
> Undoubtedly there are other ways of explaining the concept of word abstraction. The remarks above are merely intended to be indications of the kind of criteria that may be helpful in making this judgment.

After a reading of the instructions and the answering of any questions raised about procedure, each judge worked individually and at his own rate. A five-point equal-appearing-intervals scale was employed, with *1* standing for least abstract and *5* for most abstract. Judging was completed in a single session which varied for individuals from one-half hour to an hour.

Results

Reliability. Median scale values of abstraction were computed for each of 572 words from the responses of 35 judges, employing a five-point equal-appearing-intervals scale. The semi-interquartile range, *Q*, which measures the dispersion of judgments, was obtained for each scale value. The mean *Q* value is 0.64. The range of *Q* values is from .20 to 1.24, and the standard deviation is .14. Only seven of the 572 *Q* values are larger than 1.00 and only 17 larger than .90. Thus, insofar as a measure of the dispersion of judgments is also a measure of reliability, the obtained scale values appear to be satisfactorily reliable.

Pearson's correlation procedure was employed for further evaluation of the reliability of the obtained scale values. Two sets of mean scale values were computed for a sample of 90 words, the words which comprised the experimental material in the study reported by Siegel (*14*). Each of the two sets of scale values was computed from the responses of 17 judges, randomly assigned for this purpose from the entire group of 35 judges (the responses of one randomly-selected judge were omitted to provide an equal number in each group). The obtained r for estimating the strength of relationship between the two sets of scale values is .91. The size of the obtained r indicates high agreement between the two sets of values and provides strong evidence that the method of scaling results in satisfactorily reliable measures of abstraction of single words.

Median Scale Values. The particular method of computation employed

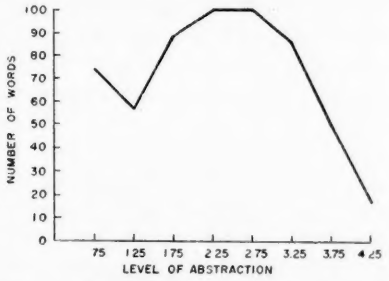

FIGURE 1. Distribution of level of abstraction scale values for 572 words. Values given along the abscissa are midpoints of degree of abstraction intervals.

allows a possible range of obtained scale values from .50 to 4.50. The mean of the obtained scale values is 2.50, which is at the midpoint of the possible range. The obtained range is from .53 to 4.43, and the standard deviation is .95.

The distribution of the obtained values, as shown in Figure 1, is essentially symmetrical and unimodal. Although there is no evidence of the true form of the distribution for the population from which the obtained 572 values are a sample, it seems reasonable to expect it to be both unimodal and symmetrical. The method employed for scaling the level of abstraction of words may thus be assumed to be useful for experimental purposes.

An end effect, that is, a piling up of words at the extremes of the scale, was anticipated as a result typical of the method of equal-appearing intervals. Such was not the case, however. The observed peaking at the extreme lower end of the scale is to be accounted for by the disproportionately high number of nouns, which, in general, were judged to be low in level of abstraction.

Figure 2 presents distributions, in percentages, of level of abstraction scale values for nouns, verbs, and adjectives, each classified as to length, 'short' and 'long': 94 short nouns, 107 long nouns, 98 short verbs, 103 long verbs, 75 short adjectives, and 95 long adjectives. It can be seen that the distributions of nouns are markedly skewed, with a preponderance of low-scale-value words. The distributions of verbs are more nearly symmetrical, while the distributions of adjectives are markedly skewed, with a preponderance of high-scale-value words. Average scale values increase in this order: short nouns, long nouns, short verbs, long verbs, short adjectives, and long adjectives. The

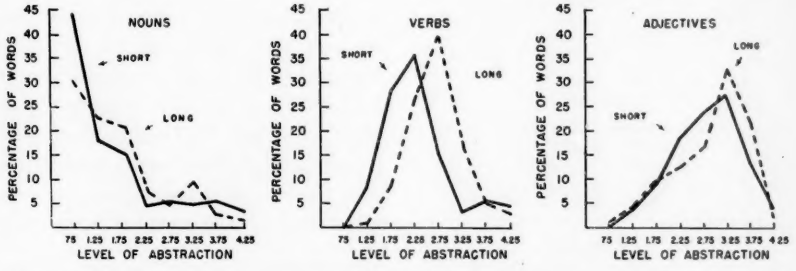

FIGURE 2. Distributions, in percentages, of level of abstraction scale values for nouns, verbs, and adjectives, each classified as to length, 'short' and 'long.' Values given along the abscissas are midpoints of degree of abstraction intervals.

entire range of scale values is represented among the short nouns, long nouns, and long adjectives; in the classifications of short verbs, long verbs, and short adjectives the entire range is represented with the exception of the lowest level of abstraction.

For the purposes of the investigation of which this scaling experiment was a part, three levels of abstraction were defined. Scale value limits were established to correspond to 'low abstraction' (Level I), 'medium abstraction' (Level II), and 'high abstraction' (Level III). Of the total list of 572 words, the 143, or 25%, which received the lowest median scale values were designated as constituting Level I; the 143, or 25%, which received the highest scale values were designated as constituting Level III; and another 143 words, or the middle 25%, were designated as constituting Level II. Specifically, words falling between the scale values of .50 and 1.63 were considered of low abstraction level; words falling between values of 2.41 and 2.88, of medium abstraction level; and words falling between the values of 3.10 and 4.50, of high abstraction level. Table 1 presents examples of words[1] representing the three parts of speech at each of the three levels of abstraction.

TABLE 1. Examples of words representing the three parts of speech at each of three levels of abstraction as scaled by 35 judges.

Levels	*Nouns*	*Verbs*	*Adjectives*
I ('low')	ant	dig	bald
	banana	fasten	chilly
	haystack	hum	sick
	oat	jump	skinny
	toe	pour	wooden
II ('medium')	business	arrange	athletic
	country	borrow	central
	family	explore	frisky
	game	join	hard
	news	keep	long
III ('high')	art	am	cute
	fame	become	dreadful
	heaven	know	honest
	liberty	seem	mere
	problem	succeed	nice

It should be acknowledged that the procedures outlined above constitute only one of several approaches to the problem of specifying abstraction level, and this approach may relate only imperfectly to Goldstein's concept of word abstraction. Goldstein, however, has provided no alternative method of quantifying his concept of word abstraction. Further study of this parameter of language appears warranted in order to refine research procedures intended to specify the language problems of dysphasic subjects and to develop more comprehensive theories of the learning and decay of language.

Summary

Median scale values of level of abstraction were obtained for each of 572 words (nouns, verbs, adjectives) from the responses of 35 judges who rated the words on a five-point equal-appearing-intervals scale. The distribution of scale values was essentially unimodal and symmetrical. Statistical analysis provided evidence that the abstraction level of single words can be reliably scaled. It is suggested that this pool of words of scaled abstraction level (low, medium, high) provides a tool for use in future studies of the parameters of language which may op-

[1]The total list of 572 words, median scale values, and *Q* values are on file at the Speech Clinic, University of Iowa.

erate differentially in language learning or in such impairments of language as dysphasia.

References

1. Aita, J. A., Modern considerations of the man with brain injury. *J. Neurosurg.*, 4, 1947, 240-254.
2. Atwell, C. R., Psychometric changes following lobotomy. In M. Greenblatt, R. Arnot, and H. C. Solomon (Eds.), *Studies in Lobotomy*. New York: Grune and Stratton, 1950.
3. Brown, I., Abstract and concrete behavior of dysphasic patients and normal subjects. *J. Speech Hearing Dis.*, 20, 1955, 35-42.
4. Buckingham, B. R., and Dolch, E. W., *A Combined Word List*. Boston: Ginn, 1936.
5. Eisenson, J., Aphasics: observations and tentative conclusions. *J. Speech Dis.*, 12, 1947, 290-292.
6. Eisenson, J., *Examining for Aphasia* (manual, rev. ed.). New York: Psychological Corp., 1954.
7. Flesch, R., Measuring the level of abstraction. *J. appl. Psychol.*, 34, 1950, 384-390.
8. Gates, A. I., *A Reading Vocabulary for the Primary Grades* (rev. ed.). New York: Teach. Coll., Columbia Univ., 1935.
9. Goldstein, K., *Language and Language Disturbances*. New York: Grune and Stratton, 1948.
10. Goldstein, K., and Scheerer, M., Abstract and concrete behavior. *Psychol. Monogr.*, 53, 1941 (whole no. 239).
11. Halstead, W. C., *Brain and Intelligence*. Chicago: Univ. Chicago Press, 1947.
12. Hebb, D. O., Man's frontal lobes. *Arch. Neurol. Psychiat.*, Chicago, 54, 1945, 10-24.
13. Johnson, W., *People in Quandaries*. New York: Harper, 1946.
14. Siegel, G. M., Dysphasic speech responses to visual word stimuli. *J. Speech Hearing Res.*, 2, 1959, 152-160.
15. Thurstone, L., and Chave, E., *The Measurement of Attitude*. Chicago: Univ. Chicago Press, 1929.
16. Weigl, E., On the psychology of the so-called processes of abstraction. *J. abnorm. (soc.) Psychol.*, 36, 1941, 3-33.
17. Weisenburg, T., and McBride, Katharine, *Aphasia, a Clinical and Psychological Study*. New York: Commonwealth Fund, 1935.
18. Wepman, J. M., *Recovery from Aphasia*. New York: Ronald Press, 1951.

Estimates of Intraglottal Pressure During Phonation

JAMES L. FLANAGAN

In an earlier report (*2*) a rationale was given for making quasi-steady approximations of the acoustic volume velocity in the human glottis during phonation. Within the conditions imposed by the assumptions, it was possible to deduce wave forms of glottal volume velocity from data on glottal area and subglottic pressure, and from steady-flow measurements of the nonlinear glottal resistance. Essentially the same arguments can be invoked to estimate the pressure variations within the glottal port.

As in the earlier report (*2*), the estimates derive from measurements reported by van den Berg and others (*3*) on the resistance of the glottal orifice to steady flow. To fix terms, a schematic representation of the vocal folds of an adult male is shown in Figure 1. The folds are approximately 3 mm in thickness, L. The glottal opening is approximated as a rectangular slit, 18 mm in length, *l*, and variable in width, w. The subglottic pressure (*re* atmospheric) is P_s, and the pressures at the entry and outlet of the glottis are P_1 and P_2, respectively. The 'downstream' (that is, pharyngeal) pressure is P_o, and for cases of open articulation is, in comparison to P_s, very nearly atmospheric. From empirical determinations, van den Berg and others give the resistance of the glottal orifice to steady flow as

$$R = P_s/Q = \frac{12\mu L}{lw^3} + 0.875\frac{\rho Q}{2l^2w^2}, \quad (1)$$

where Q is the volume velocity, μ the kinematic coefficient of viscosity, and ρ the density of moist air at body temperature.[1] Expression (1) is claimed to be accurate within 10% for: 0.1 mm $\leq$ w $\leq$ 2.0 mm; $P_s \leq$ 64 cm H_2O at small w; and Q $\leq$ 2000 cm^3/sec at large w.

If the flow through the orifice were ideal, lossless and steady, the distribution of pressure within the orifice would be as shown in Figure 2(a). The pressures at the entry and the outlet would be identical, and equal to the Bernoulli pressure, P_B. The Bernoulli pressure is the kinetic energy per-unit-volume possessed by the glottal flow, and is

$$P_B = \frac{\rho U^2}{2} = \frac{\rho Q^2}{2A^2}, \quad (2)$$

where U is the linear velocity of flow, and $A = lw$ is the glottal area. In actual measurements on plaster cast models of human larynges, however, van den Berg and others find that for 0.2 mm $\leq$ w $\leq$ 1.6 mm the distribution of pres-

James L. Flanagan (Sc.D., Massachusetts Institute of Technology, 1955) is Member of the Technical Staff, Bell Telephone Laboratories, Murray Hill, New Jersey.

[1] In c.g.s. units, the dimensions of R are dyne-sec/cm^5.

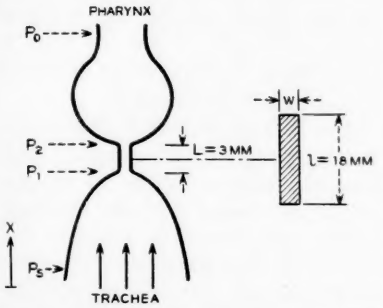

FIGURE 1. Schematic representation of vocal folds of adult male.

sure in the glottis is approximately that shown in Figure 2(b). Here,

$$P_2 = -0.5\, P_B\,, \tag{3}$$

$$\text{and } (P_S - P_1) = 1.37\, P_B\,, \tag{4}$$

for 3.6 mm^2 $\leqq$ A $\leqq$ 28.8 mm^2.

If equation (1) is solved for Q, and P_s taken as a parameter, then substitution into relation (2) yields the Bernoulli pressure as a function of glottal area, with subglottic pressure the parameter. Equation (1) is quadratic in Q, and its roots are real. Since the glottal volume flow is unipolar, and the glottal resistance nonnegative, the positive root is the appropriate value of Q. This arithmetic has been carried out for three different values of subglottic pressure, and the results are plotted in Figure 3. These data indicate the relationship between the Bernoulli pressure and the glottal area for subglottic pressures of 4, 8 and 16 cm H_2O. The dashed portions of the curves extend beyond the range of areas for which (1) is valid within an accuracy of 10%.

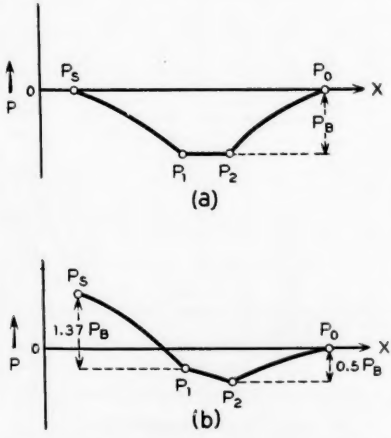

FIGURE 2. Distribution of pressure within the glottal orifice; after van den Berg and others (3): (a) ideal, (b) actual.

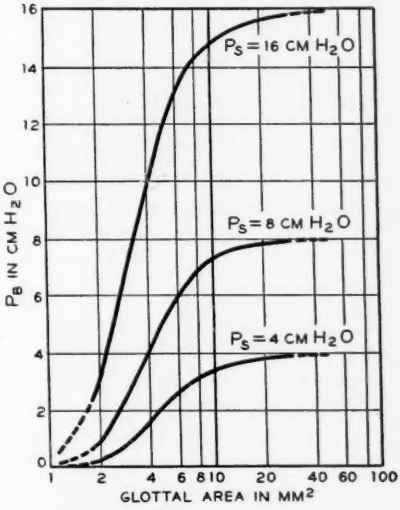

FIGURE 3. Variation of Bernoulli pressure, P_B, with glottal area. Computed formally from relations in the text.

If the subglottic pressure and the time function of glottal area are given, it is possible, therefore, to deduce from (3) and (4) quasi-steady approximations for the pressures at the entry and

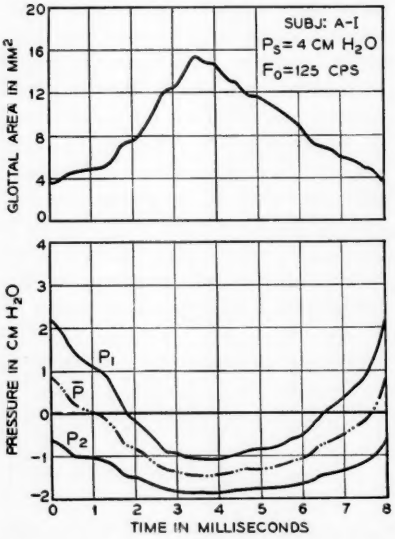

FIGURE 4. Single periods of the glottal area and pressure functions for an adult male phonating /ɑ/ at low intensity.

exit of the glottal port. To illustrate this point, glottal area functions and subglottic pressures for two conditions of phonation have been taken from the previously mentioned earlier report (2), and P_1 and P_2 have been computed for these data. Single periods of the area and pressure functions are shown for the two cases in Figures 4 and 5, respectively.

In Figure 4 the data are for an adult male phonating the vowel /ɑ/ at liminal intensity (P_s=4 cm H_2O), and at a fundamental frequency of 125 cps. For this condition complete closure of the glottis never occurs. Figure 5 is for the same person producing /ɑ/ at the greatest intensity consistent with good vowel quality (P_s=8 cm H_2O), and at a fundamental frequency of 125 cps. The dashed portions of the P_1 and P_2 curves extend beyond the range of areas for which (3) and (4) hold. In both figures the mean value of P_1 and P_2 is plotted and labeled $\bar{P}$.

The data show that P_1 can assume values ranging from the subglottic pressure at complete closure to negative values determined by the maximum opening and subglottic pressure. The pressure P_2 is always negative, and is essentially zero (or atmospheric) for complete closure. The fact that the pressures at the entry and outlet of the glottis can be of different sign suggests the possibility of a phase difference in the displacements of the lower and upper edges of the vocal folds. That is, the folds might be blown apart at the lower edges while the upper edges are being drawn together. The mode of vibration is, of course, dependent upon

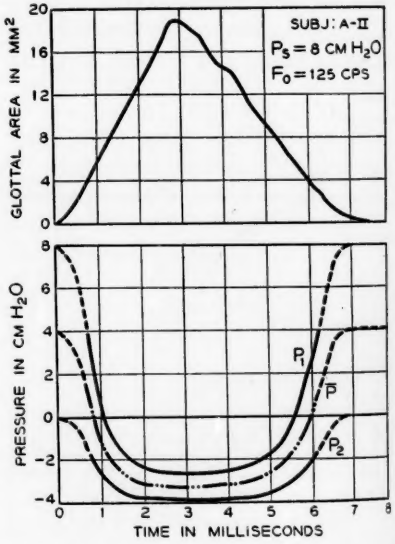

FIGURE 5. Single periods of the glottal area and pressure functions for an adult male phonating /ɑ/ at high intensity.

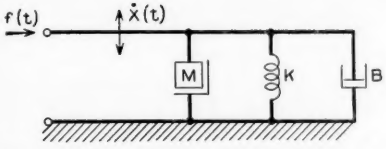

FIGURE 6. Simple mechanical oscillator composed of a mass, M, a spring, K, and a viscous resistance, B.

the distributed mass and stiffness constants of the folds, and often appears very complex. In the Bell Laboratories high-speed film of the vocal cords (*1*), however, a phase difference in displacement, such as just mentioned, seems to be quite evident for low-pitched voices in the chest register. At higher pitches, where the stiffness of the folds is greater and the mass less, the upper and lower edges appear to move in-phase.

The pressure data in Figures 4 and 5 also support observations made on the so-called breathy attack which also can be seen in the Bell Laboratories film. For this condition, vibration starts with the folds in the open phase and they are initially drawn together. If phonation is begun at a relatively large glottal area (for example, at T=4 or 5 msec in Figures 4 or 5) the mean pressure in the glottis is negative, and the initial displacement of the folds is to diminish glottal area.

An extremely crude analogy can be drawn between the mechanical vibration of the vocal folds and that of a simple mass-spring-dashpot arrangement illustrated in Figure 6. If, under conditions of low pitch and medium intensity, the displacement of the folds is approximated very roughly as

$$x(t) = k\left(1 - \cos\frac{2\pi}{T}t\right), \qquad (5)$$

where T is the fundamental vocal period, and k a scale factor, then the forcing function necessary to cause the simple mechanical circuit to execute this displacement is

$$f(t) = k\left\{\left[\left(\frac{2\pi}{T}\right)^2 M - K\right]\cos\frac{2\pi}{T}t + \left(\frac{2\pi}{T}\right)B\sin\frac{2\pi}{T}t + K\right\}. \qquad (6)$$

Wegel (*4*) has found that the natural frequency of vibration of the folds is less than the frequency produced during phonation. Consequently, sustained vibration is mass-controlled and the term $\cos\frac{2\pi}{T}t$ in (6) will predominate.

Inspection of Figures 4 and 5 shows that the intraglottal pressure functions exhibit roughly this same time relationship to their displacement (or area) functions.

Summary

From data previously reported, estimates are made of the pressure variations during phonation at the entry and outlet of the glottal port. The estimates derive from a previously advanced rationale for making quasi-steady approximations of glottal volume velocities and pressures, and are founded upon steady-flow measurements of the nonlinear glottal resistance.

Calculations for two conditions of phonation illustrate that the entry and outlet pressures can be of opposite sign, and suggest the possibility of a phase difference in the motions of the lower and upper edges of the vocal folds. In the same vein, the associated function of mean intraglottal pressure supports the observations made on the so-called

breathy attack. A crude analogy is drawn between the force and displacement relationships for the vocal folds and those of a simple, mass-controlled oscillator.

References

1. Farnsworth, D. W., High-speed motion pictures of the human vocal cords. *Bell Lab. Rec.*, 18, 1940, 203-208.
2. Flanagan, J. L., Some properties of the glottal sound source. *J. Speech Hearing Res.*, 1, 1958, 99-116.
3. van den Berg, Jw., Zantema, J. T., and Doornenbal, Jr., P., On the air resistance and the Bernoulli effect of the human larynx. *J. acoust. Soc. Amer.*, 29, 1957, 626-631.
4. Wegel, R. L., Theory of vibration of the larynx. *Bell Syst. tech. J.*, 9, 1930. 207-227.

Vowel Formant Measurements

GORDON E. PETERSON

The development of communication theory has aided considerably in defining the vocal mechanism as the primary information source in speech (*8*). Speech is structured according to the capabilities and the limitations of this mechanism. It is easily demonstrated that the resulting acoustical signal does not have an entirely unique correspondence to the positions and movements of the speech mechanism. Further, the auditory perception of speech involves a certain distortion of the acoustical signal, in which every listener applies his own aural analysis and linguistic prejudices.

Phonetic Analysis

Speech signals may be classified according to various types of direct visual and instrumental observations on the physiological mechanism of speech production. Observations on the actions of the human larynx with a high speed motion picture camera (*16*), for example, provide a basic means of categorizing laryngeal structure and function. In fact, the failure to start with technical descriptions of speech physiology and pathology is probably the primary source of the present extensive confusion about vocal qualities. In a related manner, speech signals may be classified according to acoustical similarities, as shown, for example, in sound spectrograms.

It is also possible to classify signals according to listener judgments. Various psychophysical techniques may be employed in obtaining such judgments. Accordingly, certain speech signals may be found perceptually indistinguishable or (phonetically) equivalent in certain attributes or characteristics. While physiological and acoustical measurements can be specified and defined, it is not easy to determine just what physical parameters control listener judgments about speech signal similarities.

Thus speech can be studied as a physical signal independent of language and its semantic properties. Such studies form the essential subject matter of phonetics and are basic to functional studies of spoken language.

It is obvious from texts on physiological phonetics and from more recent publications on vocal tract analogs (*1*, *15*) that many physiological dimensions must be specified in order to

Gordon E. Peterson (Ph.D., Louisiana State University, 1939) is Professor of Speech and Electrical Engineering, Director of the Speech Research Laboratory, and Chairman of the interdisciplinary program in Communication Sciences, University of Michigan. Some of the experimentation reported here was conducted while the author was a member of the technical staff, Bell Telephone Laboratories. The manuscript was prepared while the author held a Summer Faculty Research Fellowship granted by the Horace H. Rackham School of Graduate Studies, University of Michigan.

define a vowel sound accurately. In general, however, three major categories have been employed to give a simplified description of vowels: (a) position of the hump of the tongue along the vocal tract, (b) maximum degree of constriction produced by humping the tongue, and (c) degree of lip rounding. The relationship of acoustical vowel loops to these and other parameters has been discussed previously (*9*). A sophisticated technique for reducing vowel formations to three basic parameters has been described by Stevens and House (*14*), and evaluated by House and Stevens (*4*). More recently it has become evident to the present author that in some languages at least two additional properties of vowels may be involved: (d) retroflection, and (e) velarization of front vowels.

Russell (*12*) was among the first to emphasize that there is not a unique correspondence between the physiological formations of vowels and the acoustical positions of the first few formants. Russell, in fact, emphasized the difficulties which are involved in obtaining an adequate physiological description or specification of individual vowel phonemes. More recently, the work of Joos (7) and others has stressed the simplicity and convenience of describing vowels in acoustical terms. The physiological form of vowels is probably somewhat more fundamental and normally contains a greater amount of speaker information; but in a favorable acoustical environment only a limited amount of information is lost in the transformation from physiological movements and positions to sound waves. The significant acoustical parameters of vowels appear to be few in number, and the ease with which these parameters may be observed and measured in present communication technology strongly recommends the acoustical form of the vowel signal for technical study and application.

Acoustical Parameters

Various types of multidimensional signal spaces can be conceived in which all of the physical information contained within an acoustical vowel wave may be precisely specified. In the acoustical waves of actual speech there is a relatively continuous structural change, so that any given vowel in an utterance would be represented by a trace or curve within such a multidimensional space. There is every reason to believe, however, that all of the physical details of vowels are not equally significant in their formation. Those parameters which are significant may vary, of course, from one language to another. To assume that all details of vowel stimuli are controlled equally well, however, is to assume a generation of information (in the mathematical sense) which is far greater than that of which the human organism is capable. Much of the research of the past decade in acoustic phonetics has emphasized the basic significance of the formant frequencies in vowels, and this paper is primarily concerned with problems in formant frequency measurement.

The nature of the parameters of vowel production which are of primary significance may be anticipated from a consideration of the vocal mechanism. The vocal tract is essentially an acoustical tube. In vowel formations the laryngeal generator is lo-

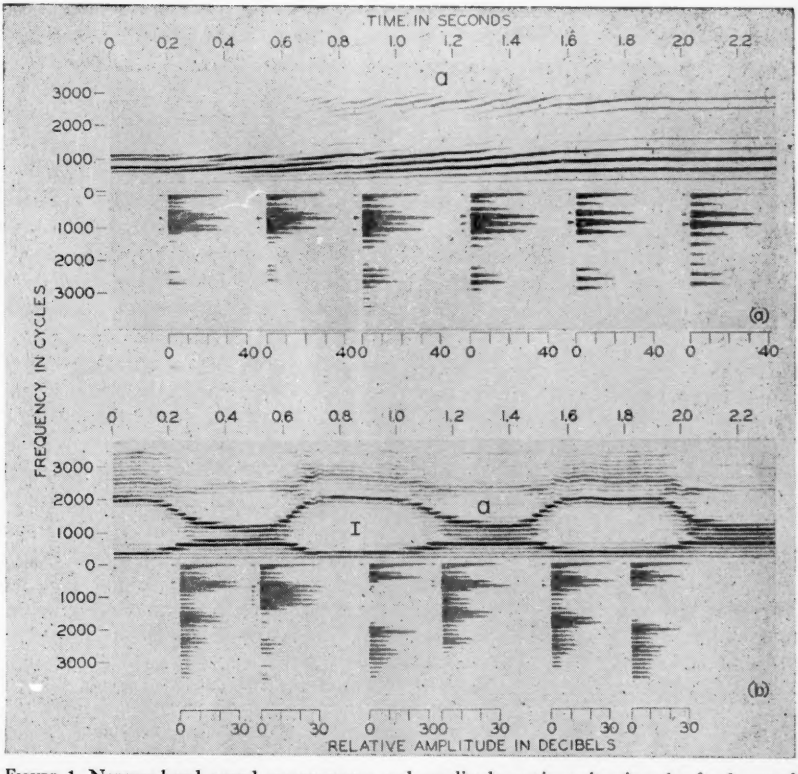

FIGURE 1. Narrow-band sound spectrograms and amplitude sections showing the fundamental voice frequency and the vowel formants. The upper picture is for a sustained vowel with rising fundamental frequency. The lower picture is for a vowel sequence with constant fundamental voice frequency.

cated at one end and sound waves are radiated from the other. The generator in turn is driven from the trachea, which serves as a supply tube. Such an acoustical system may be considered as a sequence of coupled resonance elements, with resonance peaks in the frequency-response characteristic. The parameters which may have basic importance in such a system are the fundamental driving frequency (and the input spectrum) of the generator and the frequencies, damping constants, and relative amplitudes of the resonances. Other parameters such as antiresonance frequencies may become important when the system is made more complicated, as by the addition of the nasal passageways in vowel formation.

Instrumentation

With the current apparatus of speech analysis, the various parameters indicated above may be isolated without difficulty in many vowel signals. How-

ever, there are certain circumstances under which the measurement of these parameters becomes difficult, and it is in these cases that particular attention to instrumentation is required.

The manner in which the fundamental frequency of the voice may be derived by means of the sound spectrograph is shown in Figure 1. The distinction between the voice frequency and the resonance or formant frequencies is indicated in this figure. These analyses were made with the narrow filter condition of the instrument. In the upper picture is shown a narrow filter spectrogram for a vowel with a continuously rising fundamental voice frequency. At the left of the picture, both the continuous narrow-band analysis and the amplitude sections show that there were two major regions of resonance in the general area of 1000 cps for this vowel. As voice frequency rises it becomes increasingly difficult in the figure to discriminate the two resonances; and for the higher voice frequencies there is the suggestion of just one broad resonance for the vowel.

For the lower picture of Figure 1 the fundamental frequency of the voice was held as constant as possible and the vocal cavities were altered to produce a sequence of two successive vowel positions. It will be noted that the harmonics remain essentially horizontal and that the resonance frequencies or formants vary according to the vowel.

When a relatively wide filter is employed in the analysis, the picture which results is illustrated in the upper spectrogram of Figure 2. The harmonics have now been combined to form broad bar patterns. This spectrogram is for the digits 'one, two, three.' Huggins (*5*) has modified a sound spectrograph to include an analyzing filter

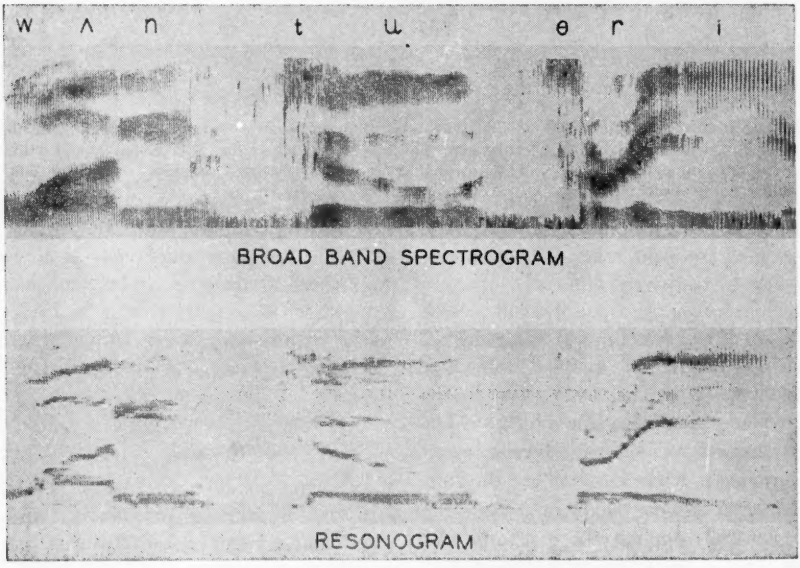

FIGURE 2. Broad-band sound spectrogram and resonogram of the same utterance.

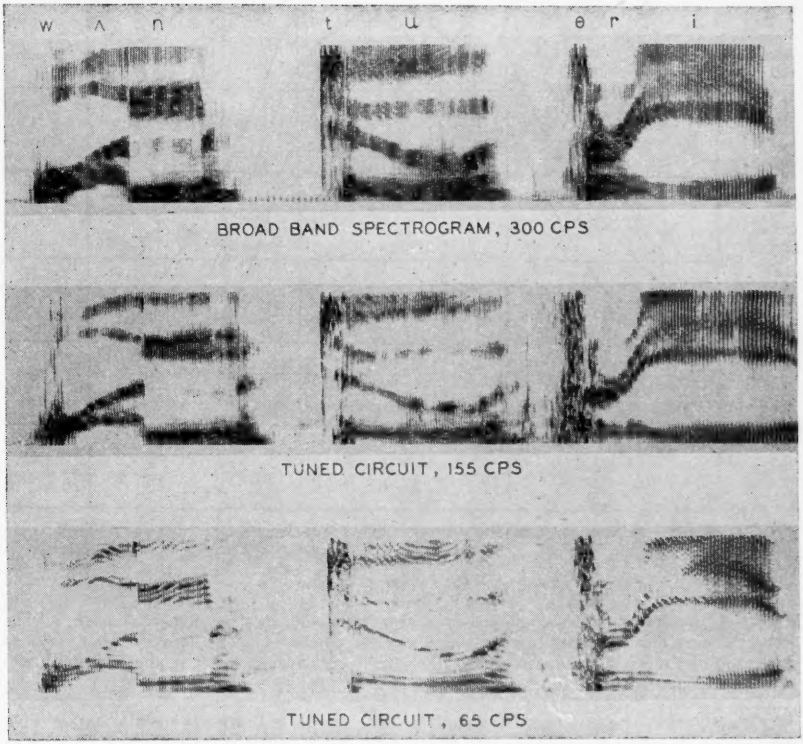

FIGURE 3. Sound spectrogram of an utterance anaylzed with the conventional wide filter and with two other tuned-circuit characteristics as shown in Figure 4.

which operates on a phase principle. Under certain circumstances such a system will display a formant with a relatively narrow line. Essentially, the lines correspond to the peaks of the envelopes of the amplitude sections shown in Figure 1. The instrument employed in marking the upper patterns of Figure 2 was modified to include such a phase filter, and an analysis of the same utterance shown in the upper spectrogram is shown with a peak or phase filter analysis in the lower picture of Figure 2.

Another, probably equally effective method of obtaining increased definition of the central formant frequency is illustrated in Figure 3. In the upper pattern of this figure is shown the conventional spectrogram of the digits 'one, two, three,' as made on the type of device employed in marking the pictures of the book *Visible Speech* (*10*). The lower two patterns show spectrograms which were obtained with simple tuned-circuit filters whose sides were sharply attenuated by the steeper slopes of the regular broad-band filter. In the instrument with which the spectrograms of Figure 3 were obtained, there

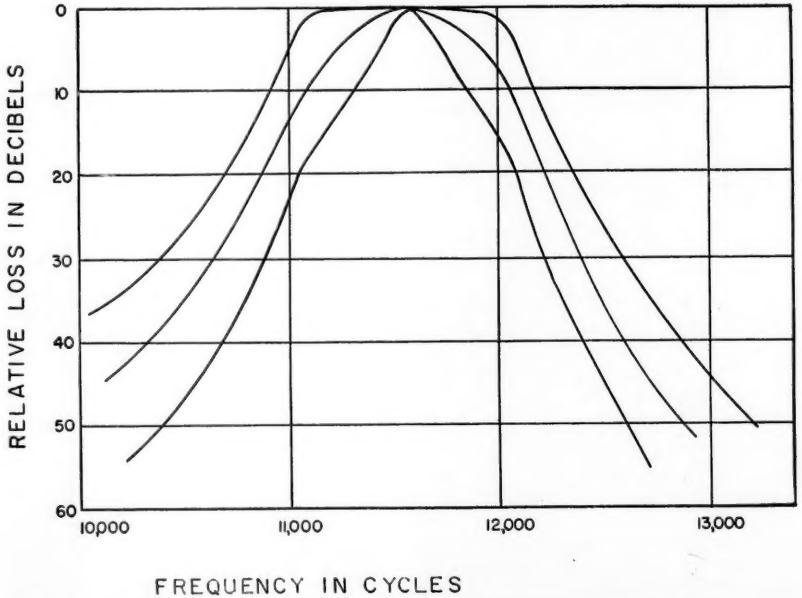

FIGURE 4. Analyzing filters employed in constructing the spectrograms of Figure 3. The two narrower characteristics are for tuned circuits combined with the wide filter. Due to the speed-up factor in the spectrograph, the effective analyzing width is approximately one-third that indicated.

is a speed-up in the analysis by a factor of approximately three. The actual filter is thus three times the width of its effective value. Frequency characteristics for the three filters employed in the analyses of Figure 3 are shown in Figure 4.

The patterns shown in Figures 2 and 3 are for a male voice with a relatively low fundamental frequency. Considerably more difficulty arises in formant display for voices with a higher fundamental voice frequency, as illustrated in the upper right-hand portion of Figure 1. In order to illustrate the effect of fundamental frequency when various analyzers are employed, simple signals were generated with the system diagrammed in Figure 5. The frequency of the oscillator controlled the frequency of the pulse generator, and the pulse generator in turn excited the tuned circuit whose frequency characteristic is shown in the graph in the upper part of the figure.

The degree of success with which the resonant peak of the tuned circuit was detected by various types of analyzers is indicated in Figure 6. The fundamental pulse frequency was varied throughout a considerable range for each of the spectrograms. In A, a resonogram was constructed with the same analyzing circuit as that employed in making Figure 2. The fundamental frequency driving the tuned circuit was varied from a low frequency, which permitted easy detection of the

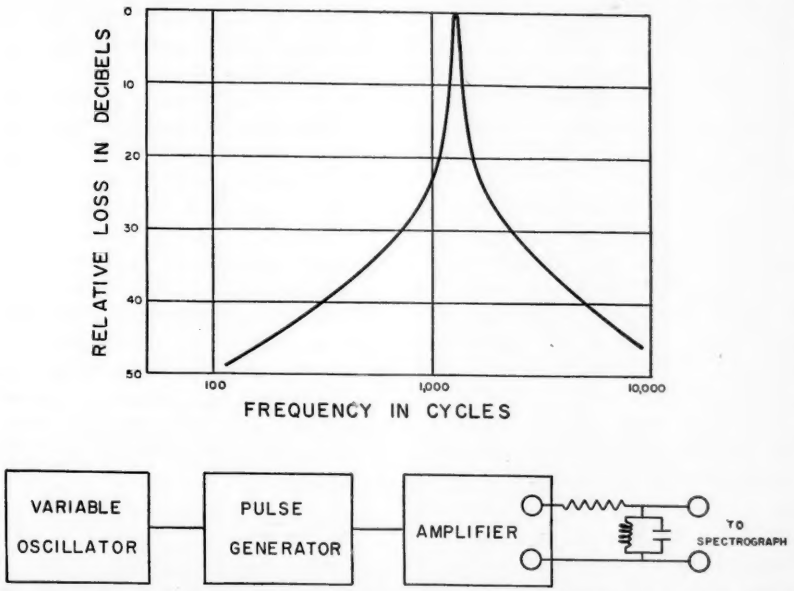

FIGURE 5. The tuned circuit whose frequency characteristic is shown in the upper figure was excited by the pulse circuit indicated in the lower figure. The repetition rate was controlled by the oscillator.

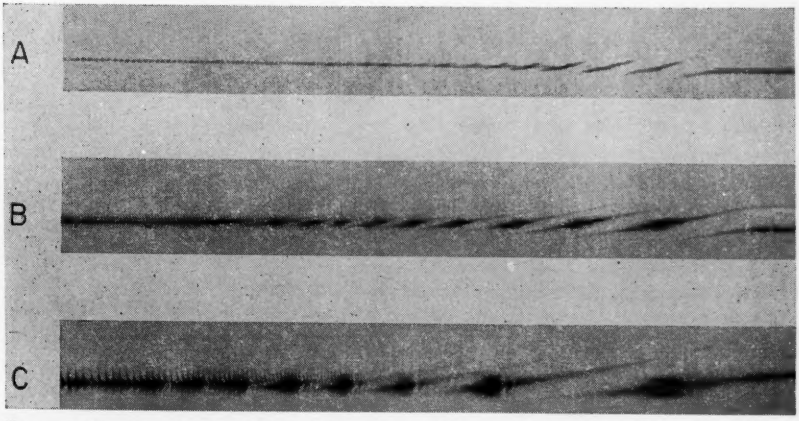

FIGURE 6. Spectrographic analyses of the output of the tuned circuit of Figure 5, when excited by a pulse of increasing repetition rate. Frequency extends vertically and increase in pulse repetition rate increases with time from left to right. A is resonogram; B, analysis with narrow tuned circuit shown in Figure 4; C, analysis with intermediate tuned circuit shown in Figure 4.

peak of the resonance, through to a point where the second harmonic of the excitation frequency was marked by the analyzer. At the right in the upper pattern in Figure 6, the correct frequency of the resonance is shown by the hand-drawn line. In B and C the analyzing tuned circuits whose responses are shown in Figure 4 were employed. In B the frequency of the input signal terminated in a value such that the second and third harmonics are displayed. In C the fundamental frequency of the signal terminated in a value within the range of the analyzer.

In such signal analyses, three primary factors are involved: the generator, the signal system which is excited by the generator, and the analyzer. Figure 6 emphasizes the fact that with current filtering techniques the success with which resonance peaks in the signal system may be detected depends very much upon the response characteristic (or transient time) of the analyzing filter relative to the recurrent rate at which the system is excited.

The following parameters are of primary interest: F_o, the fundamental frequency of excitation by the generator; F_n, the nth response (resonance) frequency of the signal system; ΔF_n, the band width in cycles per second of F_n at the 3 db points; and ΔW, the band width in cycles per second of the analyzing filter at the 3 db points.

The relationship between ΔW and F_o is probably the most important practical consideration in the spectrographic analysis of speech. When ΔW is appreciably less than F_o, the output of the analyzing filter corresponds to the amplitudes of the individual harmonics within the signal. In this case little energy will appear at the output of the analyzer at the frequency of F_n, unless F_n approximates some integral multiple of F_o.

If ΔW is appreciably greater than F_o, then the individual harmonics are normally no longer indicated, and an energy peak approximating F_n appears at the output of the analyzing filter.

When F_n is located within a complicated spectrum, the envelope curve on the two sides of the resonance may be markedly asymmetrical. Thus when $\Delta W >> \Delta F_n > F_o$, the frequency at which the maximum output of the analyzing filter occurs may not correspond closely to F_n. Thus (when $\Delta F_n > F_o$) best resonance definition will generally be obtained if $\Delta W \cong \Delta F_n$.

When $\Delta F_n >> F_o$, then the bandwidth or damping constant of the resonance is rather clearly defined in the amplitude section. An example of this is shown at the left in the upper pattern of Figure 1, where F_o is relatively low. In this case ΔF_n is well depicted.

In voices with a high fundamental frequency, however, it often occurs that $\Delta F_n >> F_o$. In such cases it is obviously impossible to derive accurately the resonator band widths from measurements of the present type of narrow-band amplitude sections. This is illustrated by the right-hand portion of the upper picture of Figure 1 where ΔF_n is largely indeterminate. Since this portion of the pattern represents essentially the same settings of the vocal cavities, as at the left of the spectrogram, there would appear to be no justification for assuming that ΔF_n has changed markedly.

In the analysis of vowels, a special difficulty occurs when two resonances approach in frequency. As indicated

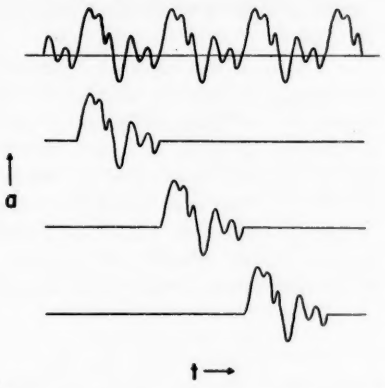

FIGURE 7. A continuous vowel wave and individual fundamental voice frequency periods derived from it for separate spectral analyses.

above, if the frequencies of the individual resonances are to be derived it is necessary to have $\Delta W > F_o$; and, for accurate specification of F_n, it is desirable that $\Delta W \cong \Delta F_n >> F_o$. An extreme case occurs when $(F_{n+1} - F_n) \leqq F_o$. If $\Delta W < F_o$, then the individual harmonics appear in the analysis; and if $\Delta W > F_o$, only one broad region of energy concentration may be observable. Thus, the difficulty is substantial whether ΔW is greater than or less than ΔF_n. The problem is particularly evident in vowels where F_2 approaches the value for F_1.

The above described inherent limitations in present techniques of spectrographic analyses can be considerably reduced by analyzing each fundamental voice period separately. With this method each period is separated out in the time domain and zero input amplitude is maintained beyond its boundaries, as illustrated in Figure 7. A discussion of the mathematical basis and possible instrumental procedures for individual pitch period analysis, however, are beyond the scope of the present paper.

The difficulties of formant detection are somewhat less severe for actual speech than for the analysis of sustained vowels. In normal speech the voice fundamental and the resonances are in almost continuous and relatively independent movement within the frequency domain. As resonances and harmonics cross, both amplitude and phase effects may be of value in observing the resonance peaks within the dynamic patterns. In general, however, the accuracy with which shifting resonance frequencies can be determined depends considerably upon the transient times involved. Normally, F_n cannot be accurately defined in appreciably less than the build-up time of the resonance; this time is approximately $1/\Delta F_n$ (*3*). When both F_o and F_n vary, there is a physical indeterminancy which depends upon their joint rates of movement in the frequency domain. The auditory perception of such dynamic acoustical patterns is very much a matter for future investigation.

Formant Measurements

Various methods of formant frequency evaluation have been suggested previously. Most of the procedures are based upon the measurement of harmonic amplitudes obtained with relatively narrow analyzing filters.

Amplitude Weighting. One method involves a weighting of the individual harmonic components according to their amplitudes within the formant (*11*). In such a weighting procedure the formant value is normally shifted toward the low frequency side of the

resonance peak. This results from the fact that the vocal cavities are of a low-pass nature in their frequency characteristic, so that the lower frequencies within the vowel spectrum generally have somewhat higher amplitudes. The magnitude of the shift generally increases as the range of frequencies considered in the formant evaluation is increased.

Spectrum Transformation. If formant measurement is to be based upon the harmonic structure of the formant, it might seem desirable to take into account the non-linear frequency and amplitude characteristics of auditory sensation. If psychophysical functions for the response of the human observer to pure tones are applied to vowel spectra, harmonic frequencies and amplitudes are converted to mels and sones. Masking effects also may be taken into account in such spectrum transformation. Certainly at this stage of psychophysical knowledge, however, caution must be exercised in the application of scales based upon elementary stimuli, such as pure tones, to the evaluation of complex signals.

When formant peaks are derived from such psychophysical vowel plots, the formant values tend to be shifted above the peak frequencies observed in the undistorted spectra. It is primarily the lower formants which are shifted, since the ear decreases rapidly in sensitivity in the low frequency region. The shifts are thus in the direction opposite from that indicated above for direct weighting of harmonic amplitudes.

Strongest Harmonic. A somewhat different and relatively simple procedure is to specify the formant frequency as the frequency of the harmonic of highest amplitude within the formant.

It is obvious that the above methods of formant frequency evaluation are essentially arbitrary in nature. None of them is adequate to deal with the highly variable conditions of fundamental frequency and spacing between formants which are found in actual speech. At one extreme only one harmonic may occur in the general region of two adjacent formant peaks.

Response Frequencies. In formant frequency evaluation, it is important to avoid a limited concept of vowel frequency structure in which only the narrow-band harmonic components are considered. If the vocal mechanism is considered to be the fundamental information source in speech, then measurements of the acoustical signal which most directly reflect its properties are of primary significance. The most fundamental functional data about such a resonant system are the response frequencies. The frequencies of the peaks of the resonances approximate the natural response frequencies of the system. Scripture (*13*) and more recently others (*2, 6*) have emphasized that it is these peaks which are of primary interest in formant frequency evaluation. The present author can see no justification for weighting calculations based on the asymmetry of formants, low-pass frequency characteristics of the vocal cavities, or mel and loudness functions.

Since the resonance peaks of the cavities are basic, it might be thought that some type of correction should be made for the spectrum of the input laryngeal pulse and the radiating characteristics of the lips. There is some

question, however, as to whether either of these change sufficiently rapidly as a function of frequency to have a significant influence upon the judgment of formant frequency peaks in sound spectrograms. In addition, the input laryngeal pulse has a spectrum which falls as a function of frequency, and the radiating characteristic of the vocal cavities is essentially rising in frequency. Thus these two effects which are superimposed upon the natural frequency response characteristic of the vocal cavities tend to cancel.

With present spectrographic techniques, then, it appears that the most meaningful procedure for formant measurement is the direct evaluation of the frequency positions of the peaks of the vowel envelope curve. Until more sophisticated techniques of instrumental analysis are available, there does not appear to be any simple procedure for the accurate derivation of formant frequencies when the formants are relatively close together in frequency and when the fundamental voice frequency is relatively high.

Summary

The measurement of formant frequency has been discussed according to the concept of the vocal mechanism as the basic information source in speech communication. The formant frequencies represent essential acoustical properties of the human vocal mechanism. Problems in the detection of these frequencies are reviewed, and procedures in the measurement of vowel spectra derived by currently available instrumental techniques are discussed. It is concluded that measurements of the peaks of the vowel envelope curves provide the most meaningful data regarding vowel formant frequencies.

References

1. Dunn, H. K., The calculation of vowel resonances, and an electrical vocal tract. *J. acoust. Soc. Amer.*, 22, 1950, 740-753.
2. Fant, C. G. M., *Acoustic Analysis and Synthesis of Speech with Applications to Swedish*. Royal Institute of Technology, Stockholm.
3. Batten, H. W., Jorgensen, R. A., Macnee, A. B., and Peterson, W. W., The response of a panoramic receiver to CW and pulse signals. *Proc. Inst. Radio Engrs*, 42, 1954, 948-956.
4. House, A. S., and Stevens, K. N., Auditory testing of a simplified description of vowel articulation. *J. acoust. Soc. Amer.*, 27, 1955, 882-887.
5. Huggins, W. H., A phase principle for complex-frequency analysis and its implications in auditory theory. *J. acoust. Soc. Amer.*, 24, 1952, 582-589.
6. Huggins, W. H., System-function analysis of speech sounds. *J. acoust. Soc. Amer.*, 22, 1950, 765-767.
7. Joos, M., Acoustic phonetics. *Language Monogr.*, 23, 1948.
8. Peterson, G. E., An oral communication model. *Language*, 31, 1955, 414-427.
9. Peterson, G. E., The phonetic value of vowels. *Language*, 27, 1951, 541-553.
10. Potter, R. K., Kopp, G. A., and Green, Harriet, *Visible Speech*. New York: D. Van Nostrand Co., 1947.
11. Potter, R. K., and Steinberg, J. C., Toward the specification of speech. *J. acoust. Soc. Amer.*, 22, 1950, 807-820.
12. Russell, G. O., *Speech and Voice*. New York: Macmillan Co., 1931.
13. Scripture, E. W., *Researches in Experimental Phonetics*. Washington, D. C.: The Carnegie Institution of Washington, 1906.
14. Stevens, K. N., and House, A. S., Development of a quantitative description of vowel articulation, *J. acoust. Soc. Amer.*, 27, 1955, 484-493.
15. Stevens, K. N., Kasowski, S., and Fant, C. G. M., An electrical analog of the vocal tract. *J. acoust. Soc. Amer.*, 25, 1953, 734-742.
16. Timcke, R., von Leden, H., and Moore, P. Laryngeal vibrations: measurements of the glottic wave. *Arch. Otolaryng.*, 68, 1958, 1-19.

Unconditioned Stimulus Strength and the Galvanic Skin Response

LEO G. DOERFLER

JOAN C. KRAMER

In the type of conditioning which is employed currently in galvanic skin response audiometry, it is important to use an unconditioned stimulus which uniformly produces a galvanic skin response. Most of the galvanic skin response testing of auditory acuity at the present time employs electric shock as the unconditioned stimulus. Studies by Doerfler and McClure (*1*), Goldstein and others (*4*), Grant and others (*5*), Humphreys (*6*), Meritser and Doerfler (7), and Stewart (*12*), which were concerned with such variables as reinforcement of the unconditioned stimulus, difficulty of conditioning subjects, and validity of galvanic skin response, all utilized electric shock for the unconditioned stimulus. The extensive use of electric shock both clinically and in research would seem to indicate that it has been an effective procedure. However, in order to effect further refinement of clinical procedures involving galvanic skin response and electric shock, it would seem that research consideration should be given to those factors which might influence ease of conditioning and resistance to extinction.

The intensity of the unconditioned stimulus as a factor in conditioning and resistance to extinction is an area that has been subjected to very little experimental work. Working with air puff as the unconditioned stimulus, Passey (*8*), Spence and Taylor (*9*), Spence (*10*), and Taylor (*13*), have all suggested that the intensity of the unconditioned stimulus is related to ease of conditioning and resistance to extinction, with the stronger unconditioned stimulus giving faster conditioning and stronger resistance to extinction.

In the study of testing of auditory acuity with galvanic skin response, using electric shock as the unconditioned stimulus, there has been no systematic attempt to indicate whether there is a difference in the conditioning and the resistance to extinction that is achieved with various levels of shock. Because clinicians in the field of audiology are constantly searching for bet-

Leo G. Doerfler (Ph.D., Northwestern, 1948) is Professor of Audiology, School of Medicine, University of Pittsburgh, and Joan C. Kramer (M.S., Pittsburgh, 1957) is Research Assistant, Jewish Hospital of St. Louis. A portion of this research was reported at the 1957 American Speech and Hearing Association Convention, Cincinnati.

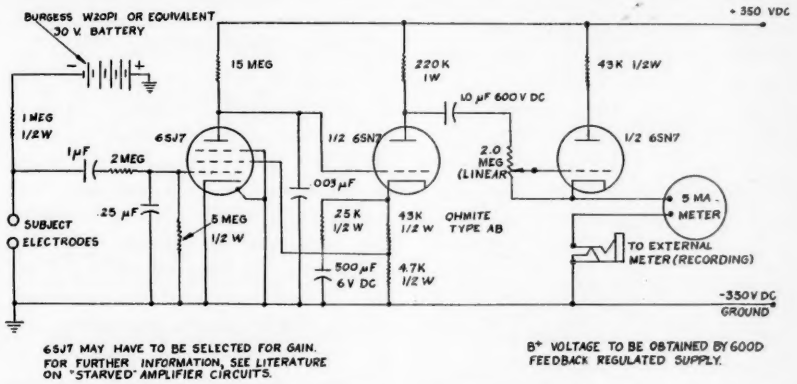

FIGURE 1. Circuit diagram for galvanic skin response amplifier.

ter defined and more systematic clinical procedures in galvanic skin response audiometry, it would appear to be of interest to investigate different levels of the unconditioned stimulus with pure tone as the conditioned stimulus and electric shock as the unconditioned stimulus.

The problem in this study was to determine the relationship between the strength of the unconditioned stimulus (shock) and acquisition of the conditioned galvanic skin response. A second problem was to determine the relationship between resistance to extinction and strength of the unconditioned stimulus.

The specific questions asked were: (a) Will acquisition of conditioning of the galvanic skin response to a pure tone differ if the strength of the unconditioned stimulus (shock) is varied in three ways: minimal, clinical, maximal? (b) Will resistance to extinction of the conditioned galvanic skin response differ if the strength of the unconditioned stimulus is varied in three ways: minimal, clinical, maximal?

Procedure

The subjects were 72 volunteers, 42 females and 30 males. These subjects ranged in age from 16 to 57 years. Restrictions imposed for selection of subjects were these two requirements: (a) response in the ear to be tested to a pure tone of 1000 cps at 15 db *re* normal threshold; (b) unfamiliarity with the principles of conditioning or galvanic skin response techniques.

The galvanic skin response apparatus used in this experiment has been described in previous studies (*1*, 7, *11*). Figure 1 shows a circuit diagram of the amplifier.

The conditioned stimulus was a pure tone produced by a commercial audiometer. The conformance of this audiometer to Bureau of Standards calibration was checked in a Type 9A coupler at the Department of Audiology, University of Pittsburgh. The 1000-cycle tone was presented at an intensity level of 30 db above normal threshold, and was always given at this level when called for by the randomized schedule. Time intervals between presentations

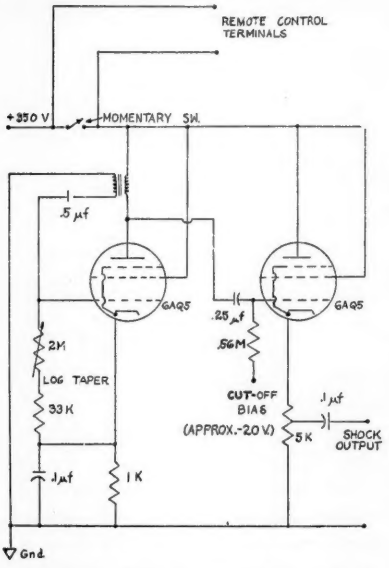

FIGURE 2. Circuit diagram for shock-producing apparatus.

(in seconds) and shock (S) or no shock (NS), in the order given, were as follows: 35, S; 40, NS; 30, NS; 40, NS; 20, S; 25, NS; 20, S; 20, NS; 35, NS; 20, S. The tone was audible for approximately 2.5 sec. A 40% randomized reinforcement schedule was employed.

The unconditioned stimulus was a pulsatile electric shock. Figure 2 shows a circuit diagram of the shock apparatus. The shock was presented for approximately .2 sec upon completion of the conditioned stimulus, when called for by the randomized schedule. Approximately .5 sec elapsed after cessation of the conditioned stimulus and the onset of the unconditioned stimulus.

The subjects were randomly assigned to one of three shock-level groups, and testing was carried on until each group consisted of 24 subjects. A total of 92 subjects was originally tested to obtain the experimental sample of 72.

Group A, minimal shock, received the least amount of shock necessary to produce a detectable skin response. This was done by setting the shock potentiometer at its lowest level and increasing it in dial units of five until two acceptable galvanic skin responses occurred at the same dial reading. Once the minimal shock level was established, the dial setting was not altered throughout the conditioning schedule.

Group B, clinical shock, was instructed to indicate to the experimenter when the shock was felt. At this point, the shock was increased five dial units, and the conditioning schedule begun. If the subject did not meet the conditioning criteria at the end of the first conditioning schedule, the shock level was increased 10 units and the second schedule begun.[1] This was the only group in which shock intensity could be altered during testing. The procedure was modeled after the standard clinical procedure during conventional galvanic skin response testing of adults at the Department of Audiology.

Group C, maximal shock, was instructed that the shock would be turned on and gradually increased. The subjects were requested and urged after each 10-unit increase in shock level to take more shock, until each subject's maximal level was reached (judged by

[1]The criteria employed were identical with those described by Stewart (*12*), Doerfler and McClure (*1*), and Meritser and Doerfler (7). They were (a) minimum GSR amplitude of 5 mm; (b) minimum GSR slope of 45°; (c) latency distance of not less than three or more than 13 mm as measured between tone onset 'pip' and leading slope of the response. Esterline-Angus chart speed was .5 in. per 5 sec.

TABLE 1. Distribution of subjects with respect to acquisition of conditioning by one, two, three, and four or more schedules. Each of three groups (N=24 for each group) received a different level of the unconditioned stimulus (shock).

Group	*Number of Schedules*			
	One	*Two*	*Three*	*Four or More*
Minimal Shock	17	2	3	2
Clinical Shock	17	4	2	1
Maximal Shock	21	1	0	2

his request to receive no more shock). Once this level was set, it was not changed during the test.

The gain control on the galvanic skin response amplifier was set at a level judged by the experimenter to be adequate to produce readable responses. This was not altered during the remainder of the conditioning and extinction procedures.

A subject was considered conditioned when he gave three consecutive responses to tones without shock, all three of which met the criteria of amplitude, latency, and slope. The procedure was set up so that at the end of the minimum conditioning schedule of 10 trials, three tones without shock were given. If the subject failed one of these responses, the schedule was immediately begun again. Once the conditioning criteria of three acceptable responses were met, he was considered conditioned and the extinction schedule started. Time intervals (in seconds) between presentations were 35, 20, 20, 25, 20, 40, 30, 40, 40, 25, 35, 25, 25, 20, 25, 35, 35, 40, 25, 25, 25, 35, 25, 25, 20. The three no-shock trials to determine conditioning were incorporated into the extinction schedule and considered to be the first three extinction trials. The extinction schedule was continued until the subject failed to give two consecutive responses meeting the criteria. Once this occurred, he was considered extinguished and the test discontinued.

The galvanic skin responses were measured with the help of a transparency, which enabled the experimenter to judge quickly and easily whether a response was acceptable.

Results and Discussion

In order to determine whether there were significant differences among the three methods of conditioning in terms of varying intensities of the unconditioned stimulus, that is, minimal, clinical, maximal, it was hypothesized that there would be no difference between each of the groups in terms of the number of subjects conditioned by one schedule or more than one schedule. Fisher's (*3*) direct method of computing probability was employed to compare the minimal with the clinical group, the minimal with the maximal group, and the clinical with the maximal group. There were no statistically significant differences between any of the groups in acquisition of conditioning. It can be seen from Table 1 that the number of subjects conditioned by one schedule in each of the groups was quite similar and that approximately the same number of subjects required more than one schedule for conditioning in each of the shock level groups.

These results would appear to suggest no relationship between intensity of the unconditioned stimulus and acquisition of conditioning as far as this study is concerned. These results differ from those reported by Passey (*8*), who indicated a relationship between strength of the unconditioned stimulus

TABLE 2. Mean, standard deviation, and range of number of trials to extinction for each of three groups, each with a different level of the unconditioned stimulus (shock).

Group	*Mean*	*S.D.*	*Range*
Minimal Shock	15.29	12.61	3–40
Clinical Shock	21.25	16.47	3–60
Maximal Shock	37.83	32.23	6–115

and acquisition of conditioning. However, if it were possible to have estimated conditioning in terms of number of trials rather than number of schedules, a similar relationship may have been revealed. The use of number of schedules as the measuring instrument was necessary in order to maintain approximately a 40% reinforcement schedule. It seems possible that differences might still exist in terms of acquisition of conditioning and intensity of the unconditioned stimulus but were not revealed because of the grossness of the measure.

In order to answer the second question raised in this study concerning the effect of the unconditioned stimulus on extinction, the mean number of trials to extinction and the variability of each of the three groups conditioned under three different strengths of shock were determined. An examination of these data, as shown in Table 2, would appear to show that the maximal group on the average resisted extinction the greatest, and that the minimal group extinguished the earliest.

TABLE 3. Results of the *t* test (N=24; *df*=23) for evaluation of differences between mean numbers of extinction trials. Comparisons are among three groups, each with a different level of the unconditioned stimulus (shock).

Groups	*Difference*	*t*	*P*
Minimal Shock Clinical Shock	5.96	1.38	N. S.*
Minimal Shock Maximal Shock	22.54	3.13	.01
Clinical Shock Maximal Shock	16.58	2.20	.05

*Not significant.

It seemed appropriate, after these differences in extinction were observed, to determine whether significant differences existed among the three groups in terms of the mean number of trials necessary for extinction. Table 3 summarizes the results of the analyses of the significance of differences, utilizing the *t* test. Because there were large differences among the standard deviations of the three groups, the usual *t* test was not employed. Cochran's (2, pp. 168, 169) formula for estimating the value of *t* was used. Clinically, the most important result to be considered was the comparison between clinical and maximal groups. The difference of 16.58 between these groups was significant at the .05 level. It would seem then, that on the average the subject remains conditioned the longest under conditions of maximal shock. However, the great variability observed might temper the clinical application of these results. There is every indication that some individuals will extinguish quite early regardless of strength of the unconditioned stimulus.

The finding that maximal shock brings about greater resistance to extinction than minimal or clinical shock is of extreme importance clinically, since resistance to extinction plays a vital role in the clinical measurement of hearing by galvanic skin response. In the course of searching for threshold, if the subject does not respond to

TABLE 4. Chi-square results for number of schedules to condition and number of trials to extinction.

Group	x^2	*df*	*P**
Minimal Shock	5.04	1	.05
Clinical Shock	1.68	1	.20
Maximal Shock	1.20	1	.30

*Because the theoretical frequencies were small, Fisher's (*3*) direct method of computing probability was also employed: P=.03, .18, and .31 for Minimal, Clinical, and Maximal Groups, respectively.

a presentation of the tone, the clinician cannot be certain whether the subject has extinguished or whether he is not hearing the tone. Chances are increased greatly that he has not extinguished when maximal shock is used in preference to minimal or clinical levels of shock.

In addition to the statistical treatment necessary to answer the two major questions posed by this study, certain comparisons were made to investigate any other relationships that might have existed. One such comparison involved determining whether those subjects who conditioned early also extinguished more rapidly within each of the three groups. It was hypothesized that within each group of those subjects extinguishing in 10 or fewer trials, there would be as many requiring one schedule to condition as there would be requiring more than one schedule to condition. The same hypothesis was made concerning those subjects extinguishing in 11 or more trials within each group. The statistic of chi square was applied. The results of this analysis are shown in Table 4. Only for the minimal group was there any evidence of a relationship between number of schedules to condition and number of trials to extinction.

TABLE 5. Mean, standard deviation, and range of number of trials from first failure to extinction. Comparisons are among three groups, each with a different level of the unconditioned stimulus (shock).

Group	*Mean*	*S.D.*	*Range*
Minimal Shock	4.71	6.35	0–20
Clinical Shock	6.79	9.15	0–29
Maximal Shock	14.54	17.80	0–59

TABLE 6. Results of the *t* test (N=24; *df*=23) for evaluation of differences between mean numbers of trials between first failure and extinction. Comparisons are among three groups, each with a different level of the unconditioned stimulus (shock).

Groups	*Mean Difference*	*t*	*P*
Minimal Shock Clinical Shock	2.08	.90	N. S.*
Minimal Shock Maximal Shock	9.83	2.49	.05
Clinical Shock Maximal Shock	7.75	1.86	.08

* Not significant.

It will be recalled that the criterion for extinction was two successive failures. It was noted that single failures often occurred that were not followed by another failure. In an effort to determine the relationship of first failure to extinction as defined in this study, an analysis of the three groups was made in terms of the number of trials between first failure and extinction. Table 5 summarizes these results. An examination of these data shows that the minimal group had the least number of acceptable responses between first failure and extinction, while the maximal group had the largest number. Also, the minimal group had the least amount of variability around the mean, indicating that it was the most reliable indication of what to expect in terms of

extinction following first failure. The general results here indicate again that extreme caution is necessary in the clinical application of these data as far as expectancy of extinction following first failure is concerned.

In Table 6 is shown the significance of differences among the three groups in terms of number of trials to extinction following first failure. Cochran's (2) formula for estimating the value of *t* was used. The difference between the minimal and clinical groups was not significant. The difference of 9.83 between the minimal group and maximal group was significant at the .05 level, indicating that the maximal group needed a significantly greater number of trials for extinction following first failure than the minimal group. The difference between the clinical and maximal groups was not significant at the .05 level. However, the *t* had a probability value of .08, indicating that only eight times in 100 would differences of this magnitude or greater arise by chance.

It became evident during the testing of the subjects receiving minimal shock that many did not acquire the conditioned response after several conditioning schedules. There were 20 of the original 92 subjects who at no time gave three consecutive acceptable responses to tone alone. Each of these subjects was taken through a minimum of six schedules (the average was more than 70 conditioning trials) before being considered unconditionable. Of these 20 subjects, 18 were in the minimal shock level group, one was in the clinical group, and one was in the maximal group. In order to ascertain whether the level of shock was the determining factor in the failure of the minimal shock group to condition, as many of this group as possible, 10 subjects, were brought back and retested using maximal shock. The 10 retested subjects were conditioned with maximal shock, nine by the end of the first schedule, that is, 90%, which is comparable to the 87.5% of the original maximal group conditioned by the first schedule. (The range of the number of trials for subjects of the retested group conditioned by the end of the first schedule was from five to 32; the number for the subject who had two schedules was 111.)

It will be recalled that one subject in the clinical and one subject in the maximal group were considered unconditionable, in terms of the criteria of this study. Both of these individuals revealed a galvanic skin response when shock was presented. However, they did not meet the criteria when the tone was given alone. It has been felt by some experimenters in the field of galvanic skin response testing that there are certain individuals who could be termed 'difficult to condition' or 'unconditionable.' Goldstein and others (4) described this situation in testing 32 normal hearing children with galvanic skin response as follows:

> Eighteen of the children were easy to condition and 14 difficult to condition according to the subjective standards used in a previous study. . . . Ease of conditioning seems to be unrelated to the intelligence quotient as measured by a performance test. Ease of conditioning also seems unrelated to sex; to age within a range of seven years and six months to twelve years and eight months; to handedness; to initial, terminal, or maximum resistance; to random fluctuations of resistance; to latency of response; and to overt responses and behavior, except for sleepiness.

It appears that regardless of the level of shock employed, there will be some

individuals who cannot be satisfactorily conditioned for purposes of galvanic skin response audiometry. However, in this sample the number of such individuals who did not condition satisfactorily appeared to be directly related to the level of the unconditioned stimulus. Those subjects who did not condition were overwhelmingly in the minimal shock group. This would tend to agree with Taylor's (*13*) finding that adaptation seems to take place more rapidly when the unconditioned stimulus is mild.

It is important to remember that even though maximal shock appears to be the best setting for maintaining conditioning, there are certain factors which could hamper its use clinically. In spite of the fact that the adults used in this study were volunteers, many of them had strong emotional reactions to the shock when it was presented at the maximal level. The decision as to whether maximal or clinical levels of shock should be used on patients in an actual clinical situation should take many things into account. If the patient is undergoing a diagnostic audiological evaluation for purposes of rehabilitation, it is possible that maximal shock could impede the establishment of rapport and retard the rehabilitative process. However, in other instances such as medico-legal cases, it may be necessary to employ maximal shock in order to insure the most accurate assessment of hearing. Still another factor that should be considered is the general health of the patient. Although there is no objective evidence either for accepting or rejecting this idea, it is possible that certain physical or mental diseases operant in a patient would contraindicate the use of maximal shock. Still another factor might be the general psychological state of the patient when he is entering the testing situation. With children, it is probably not possible to use anything above the clinical level of shock. The net result of this study would seem to indicate that there are certain advantages in the use of maximal shock, provided it is feasible to use it.

Summary

The galvanic skin response in 72 subjects with normal hearing was conditioned to a pure-tone auditory stimulus of 1000 cps using electric shock as the unconditioned stimulus. Objective criteria for designation of responses included determination of amplitude, latency, and slope. One-third (24 subjects) of the group received shock at a minimum level of intensity; one-third received shock at a clinical level of intensity; one-third received shock at a maximal level of intensity. Once conditioning was established, extinction trials were begun. All groups were extinguished in the same manner.

Results indicated that a statistically significant difference existed in number of responses to extinction among these three groups, in the direction of greater resistance to extinction by the maximal group. No statistically significant difference was found on acquisition of conditioning among the three groups.

Acknowledgment

The authors are indebted to Dr. George H. Shames, Associate Director of the Speech Clinic, University of Pittsburgh, for assistance in the statistical planning of this study.

References

1. Doerfler, L. G., and McClure, Catherine T., The measurement of hearing loss in adults by galvanic skin response, *J. Speech Hearing Dis.*, 19, 1954, 184-189.
2. Edwards, A. L., *Experimental Design in Psychological Research*. New York: Rinehart, 1950.
3. Fisher, R. A., *Statistical Methods for Research Workers*. (6th ed.) Edinburgh: Oliver and Boyd, 1936.
4. Goldstein, R., Polito-Castro, Sara B., and Daniels, Jane T., Difficulty in conditioning electrodermal responses to tone in normally hearing children, *J. Speech Hearing Dis.*, 20, 1955, 26-34.
5. Grant, D. A., Meyer, D. R., and Hake, H. W., Proportional reinforcement and extinction of the conditioned GSR, *J. gen. Psychol.*, 42, 1950, 97-101.
6. Humphreys, L. G., The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions, *J. exp. Psychol.*, 25, 1939, 141-158.
7. Meritser, C. L., and Doerfler, L. G., The conditioned galvanic skin response under two modes of reinforcement, *J. Speech Hearing Dis.*, 19, 1954, 350-359.
8. Passey, G. E., The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response, *J. exp. Psychol.*, 38, 1948, 420-428.
9. Spence, K. W., and Taylor, Janet, Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning, *J. exp. Psychol.*, 42, 1951, 183-188.
10. Spence, K. W., Learning and performance in eyelid conditioning as a function of intensity of the UCS, *J. exp. Psychol.*, 45, 1953, 57-63.
11. Stewart, K. C., A new instrument for detecting the galvanic skin response, *J. Speech Hearing Dis.*, 19, 1954, 169-173.
12. Stewart, K. C., Some basic considerations in applying the GSR technique to the measurement of auditory sensitivity, *J. Speech Hearing Dis.*, 19, 1954, 174-183.
13. Taylor, Janet, Level of conditioning and intensity of the adaptation stimulus, *J. exp. Psychol.*, 51, 1956, 127-130.

Bibliography: Delayed Auditory Feedback

RICHARD ALLEN CHASE

SAMUEL SUTTON

DAPHNE FIRST

This bibliography represents a systematic review of published literature in the area of delayed auditory feedback. Current titles are included through correspondence of the authors with investigators working in the field. Certain articles on related subjects, which were brought to the attention of the authors as a product of this correspondence, have been included in a section called 'Related Papers' presented at the end of the main bibliography.

The literature on sensory feedback may be said to go back to the nineteenth century when ideas about relationships between sensory and motor events achieved a new clarity. Early in the century Sir Charles Bell and François Magendie discovered the sensory and motor functions of the spinal roots; in 1826, Bell wrote about the 'circle of nerves' between the brain and muscles. In his *Principles of Psychology* (1890), William James speculated about the importance of sensory information to the initiation of voluntary movement and argued that the sensations subserve a 'guiding' function in the control of voluntary movement. John Dewey in 1896 argued against the unreal dualism of 'the reflex arc concept in psychology' and supported the position that stimulus and response are the functionally indistinguishable parts of a coordinated 'sensori-motor' event.

The idea that coordinated voluntary movement depends upon the utilization of sensory information by the nervous system is now generally accepted. The body of knowledge supporting this position has developed through research on different motor functions among which has been that of speech. In 1950 and 1951, Lee and Black observed a remarkable amount of disturbance in speech under conditions of delayed auditory feedback. This observation

Richard Allen Chase (A.B., University of Chicago, 1952) is a fourth-year medical student, and Research Fellow, Department of Otolaryngology and Division of Medical Psychology, Department of Psychiatry, Columbia University College of Physicians and Surgeons. Samuel Sutton (Ph.D., University of Chicago, 1955) is a Senior Research Scientist, Department of Biometrics Research, New York State Department of Mental Hygiene, and Lecturer in Psychology, Columbia University. Daphne First (A.B., University of Toronto, 1955) is a Psychological Assistant, Department of Biometrics Research, New York State Department of Mental Hygiene. This research was supported by the Department of Otolaryngology, Columbia University College of Physicians and Surgeons; the Department of Biometrics Research, New York State Department of Mental Hygiene; United States Public Health Service grants M 1541 (C1) and B 1686.

served to focus attention on the importance of sensory information to the normal control of speech. Since that time much attention has been given to this phenomenon and its applications in the field of speech and hearing. As a result there has grown a considerable literature on which the following is based.

In the compilation of the section on auditory feedback, these sources have been reviewed as indicated:

1. *Journal of Speech and Hearing Disorders*, vol. 14 (1), 1949, through vol. 23 (5), 1958.
2. *Journal of Speech and Hearing Research*, vol. 1 (1), 1958, through vol. 1 (3), 1958.
3. Graduate theses in speech and hearing research, 1949 through 1956, as reviewed in 1 and 2 above.
4. *Journal of the Acoustical Society of America*, vol. 21 (1), 1949, through vol. 30 (10), 1958.
5. *AMA Archives of Otolaryngology*, vol. 51 (1), 1950, through vol. 68 (4), 1958.
6. *Speech: The Journal of the College of Speech Therapists* (London), vol. 14 (1), 1950, through vol. 21 (2), 1957.
7. Quarterly cumulative *Index Medicus* (under: hearing, hearing disorders, hearing tests, speech, speech defects, and voice), vol. 51, 1952, through vol. 57, 1955.
8. Current *List of Medical Literature*, National Library of Medicine, U.S. Department of Health, Education, and Welfare (under: hearing, hearing disorders, hearing tests, phonetics, speech, speech disorders, and voice), vol. 20 (1), 1951, through vol. 34 (4), 1958.
9. *Psychological Abstracts* (under: audition, response processes, language and communication, and speech disorders), vol. 23 (1), 1949, through vol. 32 (2), 1958.
10. *Doctoral Dissertations*, A. H. Trotier and M. Harman, Eds., compiled for Association of Research Libraries (New York: H. W. Wilson), vol. 17, 1949-50, through vol. 22, 1954-55.
11. *Dissertation Abstracts* (University Microfilms, Inc., Ann Arbor, Mich.) (under speech), vol. 16 (1), 1956, through vol. 18 (6), 1958.

(Note: The authors would appreciate receiving information for future revisions of the bibliography. This would include new titles to be added as well as errors in the present listing.)

Auditory Feedback

1. Aldridge, R., A comparative investigation of speech under conditions of amplification, noise, and delayed auditory feedback. Unpub. master's thesis, Univ. Illinois, 1954.
2. Atkinson, C. J., Adaptation to delayed side-tone 1. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Office Naval Res., Proj. NR 142-992, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 001 064.01.12, *Joint Project Report No. 12*, 1952.
3. Atkinson, C. J., Adaptation to delayed side-tone. *J. Speech Hearing Dis.*, 1953, 18, 386-391.
4. Atkinson, C. J., Four studies on sidetone. Univ. Iowa; Signal Corps; Cont. DA 36-039 sc-42562, *Supplementary report No. 5*, 1954. (For other supplementary reports, same series, see references 14, 87, this list, and 22, *Related Papers.*)
5. Atkinson, C. J., Research study of the psycho-acoustic effects of human and artificial sidetone. Univ. Iowa; Signal Corps; Cont. DA 36-039 sc-42562, Army Proj. 3-99-12-022, *First Quarterly Report*, 1952.
6. Atkinson, C. J., *ibid*, *Second Quarterly Report*, 1953.
7. Atkinson, C. J., *ibid*, *Third Quarterly Report*, 1953.
8. Atkinson, C. J., *ibid*, *Fourth Quarterly Report*, 1953.
9. Atkinson, C. J., *ibid*, *Fifth Quarterly Report*, 1953.
10. Atkinson, C. J., *ibid*, *Sixth Quarterly Report*, 1954.
11. Atkinson, C. J., *ibid*, *Seventh Quarterly Report*, 1954.
12. Atkinson, C. J., *ibid*, *Final Report*, 1955. (See Schubert, reference 87, this list.)
13. Atkinson, C. J., Some effects on intelligibility as the sidetone level and the amount of sidetone delay are changed. *Proc. Iowa Acad. Sci.*, 1954, 61, 334-340.
14. Atkinson, C. J., Some effects on rate and on relative sound pressure level of speech produced by altering sidetone delay and sidetone level. Univ. Iowa; Signal Corps; Cont. DA 36-039 sc-42562,

Army Proj. 3-99-12-022. *Supplementary Report No. 3*, 1954.

15. Atkinson, C. J., Some influences of delayed sidetone upon intelligibility. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Off. Naval Res., U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 001 064-.01.13, *Joint Project Report, 13*, 1952, PB 107685.
16. Azzi, A., Gli impieghi della voce ritardata in fonetica (The uses of delayed voice in phonetics). *Rivista di Audiologia Practica*, 1952, 2, 75-85.
17. Azzi, A. Il Controllo auditivo della fonazione (The auditory control of phonation). *Rivista di Audiologia Practica*, 1956, 6, 65-88.
18. Azzi, A., La voce ritardata basi fisiofisiche ed applicazioni pratiche di un nuovo dispositivo electtronico per scorpire la sordità simulata (The delayed voice; psycho-physical bases and practical applications of a new electronic technique for detecting simulated deafness.) *L'Oto-Rino-Laringol. Ital.*, 1952, 21, 110-112.
19. Azzi, A., Le prove per svelare la simulazione della sordità (A test for revealing simulated deafness). *Rivista di Audiologia Practica*, 1951, 1, 22-23.
20. Azzi, A., Un nuovo metodo per svelare la simulazione della sordità (A new method for revealing the simulation of deafness). *Archivio Italiano di Otologia Rinologia e Laringologia*, 1951, 62, 329-334.
21. Azzi, A., and Invernizzi, M., La prova della voce ritardata come mezzo diagnostico nelle sordità psicogene (The 'delayed voice' test as diagnostic means in cases of psychogenic loss). *Rivista Oto-Neuro-Oftalmologica*, 1951, 26, 1-5.
22. Baccaro, P. M., A comparative study of cortical potentials of stutterers and nonstutterers under normal reading and delayed speech feedback conditions. Unpub. master's thesis, Univ. Mississippi, 1955.
23. Barnes, N. S., Delayed feedback: voice variable alterations related to intelligence levels of mentally retarded adults. Unpub. master's thesis, Purdue Univ., 1957.
24. Beaumont, J. T., and Foss, B. M., Individual differences in reacting to delayed auditory feedback. *Brit. J. Psychol.*, 1957, 48, 85-89.
25. Black, J. W., Studies of delayed sidetone. *Pubblicazoni dell' Università Cattolica del Sacro Cuore*, Nuova Serie, 1955, 49, 71-91, Società Editrice 'Vita E Pensiero,' Milano.
26. Black, J. W., Systematic research in experimental phonetics: 2. signal reception: intelligibility and side-tone. *J. Speech Hearing Dis.*, 1954, 19, 140-146.
27. Black, J. W., The effect of delayed side-tone upon vocal rate and intensity. *J. Speech Hearing Dis.*, 1951, 16, 56-60.
28. Black, J. W., The effect of room characteristics upon vocal intensity and rate. *J. acoust. Soc. Amer.*, 1950, 22, 174-176.
29. Black, J. W., The persistence of the effects of delayed side-tone. *J. Speech Hearing Dis.*, 1955, 20, 65-68.
30. Brubaker, R. S., An experimental investigation of speech disturbance as a function of the intensity of delayed auditory feedback. Unpub. doctor's dissertation, Univ. Illinois, 1952.
31. Brungard, J., An experimental investigation into the use of PGSR as an adjunct to delayed sidetone. Unpub. master's thesis, Univ. Washington, 1956.
32. Butler, R. A., and Galloway, F. T., Factoral analysis of the delayed speech feedback phenomenon. *J. acoust. Soc. Amer.*, 1957, 29, 632-635.
33. Chase, R. A., Effect of delayed auditory feedback on the repetition of speech sounds. *J. Speech Hearing Dis.*, 1958, 23, 583-590.
34. Chase, R. A., Sutton, S., and First, Daphne, A developmental study of changes in behavior under delayed auditory feedback. Communications Lab., Dept., Otolarngology, Columbia Univ., Coll. Physicians Surgeons, *Research Report*, 1957.
35. Davidson, G. D., The effect of altered external side-tone transmission time upon oral reading rate, precision of articulation and pitch variability. Unpub. doctor's dissertation, Ohio State Univ., 1955.
36. Doehring, D. G., Changes in psychophysiological responses produced by delayed speech feedback, U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Central Institute for Deaf, St. Louis; Cont. 1151(02), Proj. (Nr 146-092), U.S. Naval School Aviation Medicine, Bur. Medicine Surgery, Proj. NM 001 102 502, *Report No. 1*, 1956.
37. Doehring, D. G., and Harbold, G. J.,

The relation between speech disturbance and psychophysiological changes resulting from delayed speech feedback. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Central Institute for Deaf, St. Louis; Cont. 1151 (02), Proj. (Nr 146-092) U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 13 01 99, Subtask 1, *Report No. 5*, 1957.

38. ELKINS, E., Effects of side-tone delay on oral reading responses under conditions of binaural and monaural stimulus presentation. Unpub. master's thesis, Univ. Maryland, 1956.
39. ELLIOT, C. M., Responses of hard of hearing subjects to delayed sidetone. Unpub. master's thesis, Vanderbilt Univ. School Medicine, 1955.
40. EWERTSEN, H. W., Delayed speech test. *Acta Oto-Laryngologica*, 1955, 45, 383-387.
41. FAIRBANKS, G., Selective vocal effects of delayed auditory feedback. *J. Speech Hearing Dis.*, 1955, 20, 333-346.
42. FAIRBANKS, G., Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *J. Speech Hearing Dis.*, 1954, 19, 133-139.
43. FAIRBANKS, G., and GUTTMAN, N., Effects of delayed auditory feedback upon articulation. *J. Speech Hearing Res.*, 1958, 1 (1), 12-22.
44. FAIRBANKS, G., and JAEGER, R., A device for continuously variable time delay of headset monitoring during magnetic recording of speech. *J. Speech Hearing Dis.*, 1951, 16, 162-164.
45. FEUERMAN, H., An experimental study of anticipated difficulty of oral reading as a function of the amount of exposure to delayed auditory feedback. Unpub. master's thesis, Penn. State Univ., 1957.
46. FORS, E. W., An experimental investigation of the time delay interval which produces maximum speech disturbance during delayed auditory feedback. Unpub. master's thesis, Univ. Illinois, 1954.
47. FRY, D. B., The experimental study of speech. In A. J. Ayers and others (Eds.), *Studies in Communication:* contributed to Communication Res. Center, Univ. College, London, 154-155. London: Martin Secker and Warburg, 1955.
48. GIBBONS, E. W., and WINCHESTER, R. A., A delayed sidetone test for detecting uniaural functional deafness. *Arch. Otolaryngol.*, 1957, 66, 70-78.
49. GOLDFARB, W., and BRAUNSTEIN, P., Reactions to delayed auditory feedback in schizophrenic children. In P. H. Hoch and J. Zubin (Eds.), *Psychopathology of Communication*, 49-63. New York: Grune and Stratton, 1958.
50. GROVER, J. M., A study comparing adaptation to stuttering with adaptation to auditory delayed speech feedback. Unpub. master's thesis, Univ. Wichita, 1956.
51. GUTTMAN, N., An analysis of articulation as a function of delayed auditory feedback. Unpub. master's thesis, Univ. Illinois, 1951.
52. GUTTMAN, N., Experimental studies of the speech control system. Unpub. doctor's dissertation, Univ. Illinois, 1954.
53. HAM, R. E., Certain effects on speech of alterations in the auditory feedback of speech defectives and normals. Unpub. doctor's dissertation, Purdue Univ., 1956.
54. HANLEY, C. N., and TIFFANY, W. R., Auditory malingering and psychogenic deafness. *Arch. Otolaryngol.*, 1954, 60, 197-201.
55. HANLEY, C. N., and TIFFANY, W. R., An investigation into the use of electromechanically delayed side-tone in auditory testing. *J. Speech Hearing Dis.*, 1954, 19, 367-374.
56. HANLEY, C. N., TIFFANY, W. R., and BRUNGARD, J. M., Skin resistance changes accompanying the sidetone test for auditory malingering. *J. Speech Hearing Res.*, 1958, 1, 286-293.
57. HARBOLD, G. J., and DOEHRING, D. G., A rating scale measure of speech disturbances that accompany delayed speech feedback. U. S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Off. Naval Res. Proj. NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 180299, Subtask 1, *Report No. 71*, 1957.
58. HARFORD, E. R., The effect of loudness recruitment on delayed speech feedback. Unpub. doctor's dissertation, Northwestern Univ., 1958.
59. HERBST, I., An experimental study of the effect of room characteristics and noise upon the rate and accuracy of reading familiar material. Unpub. master's thesis, Ohio State Univ., 1951.
60. HULL, F. M., An experimental investigation of speech disturbance as a func-

tion of the frequency distortion of delayed auditory feedback. Unpub. doctor's dissertation, Univ. Illinois, 1952.

61. Huntington, D. A., An experimental investigation of the relationships between auditory abilities and the speech disturbances produced by delayed auditory feedback. Unpub. doctor's dissertation, Univ. Illinois, 1952.
62. Jordan, E. P., A comparative investigation of impromptu speaking and oral reading under conditions of delayed auditory feedback. Unpub. master's thesis, Univ. Illinois, 1952.
63. Kalmus, H., Denes, P., and Fry, D. B., Effect of delayed acoustic feed-back on some non-vocal activities. *Nature*, London, 1955, 175, 1078.
64. Kline, M. V., Guze, H., and Haggerty, A. D., An experimental study of the nature of hypnotic deafness: effects of delayed speech feedback. *J. clin. exp. Hypnosis*, 1954, 2, 145-156.
65. König, G., Sprachverzögerungstest zum nachweis der simulation (Delayed speech test as a proof of malingering). *Z. Laryngol. Rhinol. Otol.*, 1955, 34, 748-751.
66. Lee, B. S., Artificial stutter. *J. Speech Hearing Dis.*, 1951, 16, 53-55.
67. Lee, B. S., Effects of delayed speech feedback. *J. acoust. Soc. Amer.*, 1950, 22, 824-826.
68. Lee, B. S., Some effects of side-tone delay. *J. acoust. Soc. Amer.*, 1950, 22, 639-640.
69. Leith, W. R., and Pronko, N. H., Speech under stress: a study of its disintegration. *Speech Monogr.*, 1957, 24, 285-291.
70. Lerche, E., and Nessel, E., Neue beobachtungen bei reihenuntersuchungen mit verzögerter sprachrückkopplung (Lee-effekt) [Further observations in serial studies with delayed speech feedback (Lee effect)]. *Archiv für Ohren-Nasen-und Kehlkopf-Heilkunde*, 1956, 169, 505-508.
71. Marple, N. B., and Morrill, S. N., A device for the production of delayed side-tone. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola, Bur. Medicine Surgery, Res. Proj. NM 001 064.01.08, *Joint Project Report No. 8*, 1951.
72. McCroskey, Jr., R. L., Some effects of anaesthetizing the articulators under conditions of normal and delayed side-tone. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Off. Naval Res., Proj. NR 145-993, U.S. Naval School Aviation Medicine Bur. Medicine Surgery, Proj. NM 001 104 500, *Joint Project Report No. 65*, 1956.
73. Melrose, J., An experimental study of speech under conditions of delayed auditory feedback between one and five seconds. Unpub. master's thesis, Univ. Illinois, 1953.
74. Melrose, J., The temporal course of changes in the amount of vocal disturbance produced by delayed auditory feedback. Unpub. doctor's dissertation, Univ. Illinois, 1954.
75. Meyer-Eppler, W., Verzögerte rückkopplung als mittel der stimmforschung (Delayed feedback as a means of speech investigation). *Elektr. Rdsch.*, 1956, 10, 91-93,
76. Meyer-Eppler, W., and Luchsinger, R., Beobachtungen bei der verzögerten rückkopplung der sprache (Lee-effekt) [Observations on the reactions to delayed auditory feedback (Lee effect)]. *Folia Phoniatrica*, 1955, 7, 87-99.
77. Neely, K. K., The effect of oral practice in presence of different conditions of side-tone upon the rate and sound pressure level of the speech of a group of stutterers. Unpub. doctor's dissertation, Ohio State Univ., 1951.
78. Nidetz, M., An experimental study of speech under conditions of delayed auditory feedback: time delay between .01 and .10 second. Unpub. master's thesis, Univ. Illinois, 1953.
79. Oshrine, M., An analysis of relationships between measurement of reading skills, speech performances, speech attitudes and measurements of oral reading rate under conditions of delayed side-tone. Unpub. master's thesis, Univ. Maryland, 1956.
80. Peters, R. W., The effect of changes in external side-tone delay and level upon rate of oral reading of normal speakers. Unpub. doctor's dissertation, Ohio State Univ., 1953.
81. Peters, R. W., The effect of changes in side-tone delay and level upon rate of oral reading of normal speakers. *J. Speech Hearing Dis.*, 1954, 19, 483-490.
82. Pronko, N. H., and Leith, W. R., Behavior under stress: a study of its disintegration. *Psychol. Rept.*, 1956, 2, 205-222.

83. Quigley, S. P., The vocal effects of delayed auditory feedback and their relationships to other individual characteristics. Unpub. doctor's dissertation, Univ. Illinois, 1957.
84. Rabenstein, Jr., J. E., Relationship between responses to delayed auditory feedback and embedded-figures test. Unpub. master's thesis, Pennsylvania State Univ., 1958.
85. Rawnsley, A. I., and Harris, J. D., Comparative analysis of normal speech and speech with delayed side-tone by means of sound spectrograms. U. S. Naval Submarine Base, New London, Bur. Medicine Surgery, Navy Dept. Proj. NM 003 041.56.03, *Medical Research Laboratory Report No. 248, 13 (a)*, 1954.
86. Robinson, F. B., Effects of changes in the relationship between the speech and the external side-tone level on the oral reading rate of stutterers and nonstutterers. Unpub. doctor's dissertation, Ohio State Univ., 1951.
87. Schubert, E. D., Research study of the psycho-acoustic effects of human and artificial sidetone. Univ. Iowa; Army; Proj. 3-99-12-022, Cont. DA36-039, sc-42562, *Final Report*, 1954, (Atkinson series).
88. Schwartz, R. J., Vocal responses to delayed auditory feedback in congenitally blind adults. Unpub. doctor's dissertation, Purdue Univ., 1956.
89. Siegenthaler, B. M., and Brubaker, R. S., Suggested research in delayed auditory feedback. *Pennsylvania Speech Annual*, 1957, 14, 24-31.
90. Sills, S. A., Some effects of delayed side-tone on the speech variables of children. Unpub. master's thesis, Purdue Univ., 1957.
91. Spilka, B., A study of relationships existing between certain aspects of personality and some vocal effects of delayed speech feedback. Unpub. doctor's dissertation, Purdue Univ., 1952.
92. Spilka, B., Relationships between certain aspects of personality and some vocal effects of delayed speech feedback. *J. Speech Hearing Dis.*, 1954, 19, 491-503.
93. Spilka, B., Some vocal effects of different reading passages and time delays in speech feedback. *J. Speech Hearing Dis.*, 1954, 19, 37-47.
94. Spilka, B., Hanley, T. D., and Steer, M. D., Relationships between certain aspects of personality and some vocal effects of delayed speech feedback. Navy, Off. Naval Res., Spec. Devices Ctr., Specdevcen 104-2-37; Purdue Univ. Specdevcen Proj. 20-F-8, Cont. N6ori-104, *Technical Report*, 1954.
95. Spuehler, H. E., Effects and interactions of delayed side-tone and auditory flutter. Unpub. doctor's dissertation, Purdue Univ., 1956.
96. Tenaglia, G., Contribrito ad un nuovo metado per svelare la sordità a mezzo della voce ritardata (New method of discovering deafness by means of delayed speech). *Giornale di Medicina Militaire*, 1954, 104, 299-301.
97. Tiffany, W. R., and Hanley, C. N., Adaptation to delayed side-tone. *J. Speech Hearing Dis.*, 1956, 21, 164-172.
98. Tiffany, W. R., and Hanley, C. N., Delayed speech feedback as a test for auditory malingering. *Science*, 1952, 115, 59-60.
99. Tiffany, W. R., Hanley, C. N., and Sutherland, L. C., A simple mechanical adapter for variable side-tone delay. *J. Speech Hearing Dis.*, 1954, 19, 504-506.
100. White, R., The relationship between oral reading rate and the interval of delayed auditory feedback causing maximum disturbance. Unpub. master's thesis, Penn. State Univ., 1958.
101. Winchester, R. A., and Gibbons, E. W., Relative effectiveness of three modes of delayed sidetone presentation. *Arch. Otolaryngol.*, 1957, 65, 275-279.

Related Papers

1. Anderson, B. A., and Knauf, V. H., Two ways of increasing the clinical usefulness of a magnetic tape recorder. *J. Speech Hearing Dis.*, 1949, 14, 37-42.
2. Atkinson, C. J., A study of vocal responses during controlled aural stimulation. Unpub. doctor's dissertation, Ohio State Univ., 1950.
3. Atkinson, C. J., Vocal responses during controlled aural stimulation. *J. Speech Hearing Dis.*, 1952, 17, 419-426.
4. Békésy, G. v., The structure of the middle ear and the hearing of one's own voice by bone conduction. *J. acoust. Soc. Amer.*, 1949, 21, 217-232.
5. Békésy, G. v., Vibration of the head in a sound field and its role in hearing by bone conduction. *J. acoust. Soc. Amer.*, 1948, 20, 749-760.
6. Black, J. W., A compensatory effect in vocal responses to stimuli of low intensity. *J. exp. Psychol.*, 1950, 40, 396-397.

7. Black, J. W., Application of side-tone in subjective tests of microphones and headsets. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State. Univ. Res. Found.; Cont. N6ONR 22525, Off. Naval Res. Proj. Designation NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery, Proj. NM 001 064.01.20, *Joint Report No. 20*, 1954, PB 114247.
8. Black, J. W., Control of the sound pressure level of voice. U. S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Off. Naval Res., Proj. Designation NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery, Proj. NM 001 104 500.42, *Joint Report No. 42*, 1955, PB 117653.
9. Black, J. W., Some effects of auditory stimuli upon voice. *J. Aviat. Med.*, 1950, 21, 251-255 and 277.
10. Black, J. W., Some effects upon voice of hearing tones of varying intensity and frequency while reading. *Speech Monogr.*, 1950, 17, 95-98.
11. Black, J. W., The effect of noise-induced temporary deafness upon vocal intensity. *Speech Monogr.*, 1951, 18, 74-77.
12. Black, J. W., The loudness of side-tone. *Speech Monogr.*, 1954, 21, 301-305.
13. Black, J. W., The relation between message-type and vocal rate and intensity. *Speech Monogr.*, 1949, 16, 217-220.
14. Black, J. W., and Morrill, S. N., Pitch of side-tone. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22523, Off. Naval Res. Proj. Designation NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 001 064.01.31, *Joint Project Report No. 31*, 1954, PB 116530.
15. Black, J. W., and Tolhurst, G. C., Intelligibility as related to the path of airborne side-tone. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Office Naval Res. Proj. Designation NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 001 064.01.34, *Joint Project Report No. 34*, 1954.
16. Bliss, E. L., Migeon, C. J., Hardin Branch, C. H., and Samuels, L. T., Reaction of the adrenal cortex to emotional stress. *Psychosomatic Med.*, 1956, 18, 56-76 (p. 68).
17. Broadbent, D. E., Speaking and listening simultaneously. Med. Res. Council, Appl. Psychol. Res. Unit, Psychol. Lab., Cambridge, Eng., *Report No. 146/151*, 1951.
18. Broadbent, D. E., Speaking and listening simultaneously. *J. exp. Psychol.*, 1952, 43, 267-273.
19. Cherry, C., and Sayers, B. McA., Experiments upon the total inhibition of stammering by external control, and some clinical results. *J. Psychosomatic Res.*, 1956, 1, 233-246.
20. Cutler, H. M., The effects of stress on the perception of time by normal and neurotic subjects. Unpub. doctor's dissertation, Purdue Univ., 1951.
21. Dittman, H. H., The role of proprioceptive sensibilities in speech production. Unpub. doctor's dissertation, Univ. Denver, 1955.
22. Dolch, J. P., and Schubert, E. D., Study of body-conducted sidetone. Univ. Iowa; Signal Corps; Cont. DA 36-039 sc-42562, *Supplementary Report No. 6*, 1954.
23. Draegert, G. L., Relationships between voice variables and speech intelligibility in high level noise. *Speech Monogr.*, 1951, 18, 272-278.
24. Fairbanks, G., Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *J. Speech Hearing Dis.*, 1954, 19, 133-139.
25. Fletcher, H., Raff, G. M., and Parmley, F., Study of the effect of different amounts of side-tone in the telephone set. Western Electric Co., *Report No. 19412*, Case No. 120622, 1918.
26. Funkenstein, D. H., King, S. H., and Drolette, Margaret E., *Mastery of Stress*. Cambridge, Mass.: Harvard Univ. Press, 1957.
27. Gibbons, E. W., Winchester, R. A., and Krebs, D. F., The variability of oral reading rate. *J. Speech Hearing Dis.*, 1958, 23, 591-593.
28. Gibbs, C. B., The continuous regulation of skilled response by kinaesthetic feed back. *Brit. J. Psychol.*, 1954, 45, 24-39.
29. Granit, R., Systems for control of movement. Premier Congrès International des Sciences Neurologiques Bruxelles, 21-28 Juillet 1957. Extrait du volume publié à l'occasion de la première journée commune, 63-99.
30. Hammond, P. H., Merton, P. A., and Sutton, G. G., Nervous gradation of muscular çontraction. *Brit. Med. Bull.*, 1956, 12, 214-218.

31. Hanley, T. D., and Steer, M. D., Effect of level of distracting noise upon speaking rate, duration and intensity. *J. Speech Hearing Dis.*, 1949, 14, 363-368.
32. Hefferline, R. F., The role of proprioception in the control of behavior. *Trans. N.Y. Acad. Sciences*, 1958, 20, 739-764.
33. Korobow, N., Reactions to stress: a reflection of personality trait organization. *J. abnorm. soc. Psychol.*, 1955, 51, 464-468.
34. Korobow, N., The relationship between personality and reactions to audiogenic stress. Unpub. doctor's dissertation, New York Univ., 1953.
35. Lightfoot, C., and Morrill, S. N., Loudness of speaking: the effect of the intensity of side-tone upon the intensity of the speaker. U.S. Naval School Aviation Medicine Res., Naval Air Station, Pensacola; Kenyon College; Cont. N7 onr-411 T. O. I., Off. Naval Res., Proj. Designation NR 782 004, U.S. Naval School Aviation Medicine Res., Bur. Medicine Surgery Proj. NM 001 053, *Joint Project Report No. 4*, 1949, PB 115196.
36. Malloy, M. M., Investigation of the relationship between the pitch of side-tone and the frequency of the voice. Unpub. master's thesis, Ohio State Univ., 1953.
37. Peters, R. W., The effect of filtering of side-tone upon speaker intelligibility. *J. Speech Hearing Dis.*, 1955, 20, 371-375.
38. Peters, R. W., The effect of high-pass and low-pass filtering of side-tone upon speaker intelligibility. U.S. Naval School Aviation Medicine, Naval Air Station, Pensacola; Ohio State Univ. Res. Found.; Cont. N6ONR 22525, Off. Naval Res. Proj. Designation NR 145-993, U.S. Naval School Aviation Medicine, Bur. Medicine Surgery Proj. NM 001 064.01.25, *Joint Project Report No. 25*, 1954.
39. Peterson, G. E., An oral communication model. *Language*, 1955, 31, 414-427.
40. Poulton, E. C., Simultaneous and alternate listening and speaking. *J. acoust. Soc. Amer.*, 1955, 27, 1204-1207.
41. Smith, W. M., Smith, K. U., Stanley, R., and Harley, W., Analysis of performance in televised visual fields: preliminary report. *Percep. Motor Skills*, 1956, 6, 195-198.
42. Stromsta, C. P., A first approximation of the distance from vocal cords to cochlea and the transit time of bone-conducted sound from the region of the vocal cords to the region of the cochlea. Unpub. master's thesis, Ohio State Univ., 1951.
43. Verzeano, M., Time-patterns of speech in normal subjects. *J. Speech Hearing Dis.*, 1950, 15, 197-201.
44. Verzeano, M., Time-patterns of speech in normal subjects. II. *J. Speech Hearing Dis.*, 1951, 16, 346-350.
45. Winchester, R. A., and Gibbons, E. W., The effect of auditory masking upon oral reading rate. *J. Speech Hearing Dis.*, 1958, 23, 250-252.
46. Wood, K. S., A preliminary study of speech deterioration under complete binaural masking. *West. Speech*, 1950, 14, 38-40.

Acknowledgment

The authors gratefully acknowledge the generous support and assistance of Dr. Edmund P. Fowler, Jr., Dr. Joseph Zubin, Dr. Newman Guttman, Miss Gina Aversa, and Mr. Lloyd Gilden.